GENERALISED EXPONENTIALLY WEIGHTED REGRESSION

AND

DYNAMIC BAYESIAN FORECASTING MODELS

MUHAMMAD AKRAM

THESIS SUBMITTED IN FULFILMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

STATISTICS DEPARTMENT

UNIVERSITY OF WARWICK

COVENTRY CV4 7AL

ENGLAND - U.K.

1st June, 1984

In Loving Memory Of My Father:

Chaudhary Muhammad Hussain sahibzada Mian Rukan Din

To:

Sultana

Sumera

Sumrana

Khuram

And All Those To Whom I Really Love.

ACKNOWLEDGEMENT

I am indebted to Professor P J Harrison who introduced to me the concept of time series analysis and forecasting. His supervision, guidance and continuous encouragement enabled me to complete the research project. For his kind and sympathetic attitude I shall always be grateful to him.

My thanks are due to Dr. M West and other staff members of the Statistics Department with whom I often had useful discussions.

I would also express gratitude to the examiners Dr. K D Stoodley and E Godolphin for their valuable comments.

I am very grateful to my mother, wife Sultana and children Sumera, Sumrana and Khuram who suffered a lot during the period of my study and never complained about their sufferings.

ABSTRACT

In this dissertation a new Forecasting Methodology, called Generalised Exponentially Weighted Regression (G.E.W.R) is presented. This Methodology is based on Linear Filtering using an Exponentially Weighted System and a Bayesian formulation developed. It is particularly designed to analyse discrete time series driven by Autoregressive and Moving Average type Coloured Noise processes. In order to elaborate the theory various theorems and corollaries are given.

For the implementation of G.E.W.R. various parsimonious Bayesian Dynamic Linear Models and Normal Discount Models for Low and High Frequency Components of time series with or without Seasonality and Cyclicity are introduced.

For theoretical and computational purposes recurrence relations for the Precision and Transformation Matrices are developed.

For the unknown variance case an automatic (self-tuning) on line Bayesian Learning Procedure is introduced.

For Complex Systems a procedure to construct the State Space Models is given and, for practitioners, methods of reparameterising Dynamic Linear Models is given.

In order to demonstrate the performance of G.E.W.R. the theory is applied to various simulated data sets and real life economic and industrial time series. In all cases the Methodology not only generates one-step ahead optimum forecasts in a Minimum Mean Square Error (M.M.S.E.) sense but also provides reasonable long term forecasts.

CONTENTS

1

INTRODUCTION AND PRELIMINARIES

1.1 Historical Review and Introduction

The technique of assigning weights to observations according to their relative importance is quite old. The scientists of 300 B.C. were quite aware of the advantages of such a system. They applied it in their Astronomical Studies.

In 1722 Roger Cotes estimated planetary parameters by assigning weights to the measurements. Euler in 1749 used weighted averages to study the irregular motion of Saturn and Jupiter. In 1797 James Bernoulli remarked: " It is absurd to allocate the same weight to several observations as the uniform distribution in many cases may not describe the measurement error appropriately ".

Our ancestors, clearly, seemed to be convinced by the use of a weighting system in parameter estimation, which even now is extensively being used in the analysis of time series, though in a different form. The present system in which the weights are allocated to the observations such that the maximum information is obtained from the present observation and the information from the past is discounted exponentially is due to Holt (1957). This technique has steadily been advanced by Brown (1959), Winters (1960), Muth (1961), Cox (1961), Brown-Meyer (1961), D'Esopo (1961), Dobbie (1963), Harrison (1965, 67, 71, 76), Gilchrist (1967), Godolphin-Harrison (1975), Mckenzie (1974, 76), etc. and most recently by Harrison-Akram (1983), Akram-Harrison (1983) and Ameen-Harrison (1983).

The roots of the Bayesian philosophy also lie quite deep in the remote past. It was and is natural to learn a lesson from history in order to project the future by utilizing present information or knowledge. A systematic approach to this philosophy was first given by (probably) Thomas Bayes in the seventeen century (a controversial issue - Stigler (1983)). No matter who developed it!. This useful and realistic approach which has now become an important ingredient of Statistics is known as Bayesian Statistics. This is now commonly being used in probability theory, time series analysis, statistical inference and other related areas. There are many notable works in this area such as those by De Finetti (1964), Fukao (1965), De Groot (1970), Lindley (1975), Harrison-Stevens (1971, 76), Lindley-Smith (1972), Peterka (1981), West (1981, 82), etc.

Compared to the above, the Control theory approach seems young. The earliest examples on optimal control of discrete time system may be found in Kalman-Koepcke (1958). However a more comprehensive approach was given by Kalman (1960) for discrete time systems. This is often known as the Kalman Filter for Linear Discrete Systems. Harrison-Stevens (1971, 76) introduced Bayesian forecasting to time series analysts and forecasting practitioners and, in the Normal case, their recurrence formulae are the Kalman Filter equations. They presented a wide class of Bayesian Dynamic Linear Models, which are now being extensively used in theory and practice.

In this dissertation the G.E.W.R. Methodology presented is developed by utilizing the useful ideas of all three approaches, along with the concept of coloured noise processes characterised by Autoregressive and Moving Average processes.

The theory developed is made capable of coping with time series driven by a coloured noise process. The Methodology presented is for discrete time series.

1.2 Object of the Research

The main aim of the present research was two fold:

1) To extend concept of exponential smoothing or discounting to the case of discrete stationary time series having an ARMA type coloured noise component, as it is unrealistic to assume that the noise component of a time series is always a white noise process. In real life generally, economic, social, industrial and other time oriented scientific data series show the characterstics of coloured noise.

2) To develop a joint modelling scheme where the Low, Medium and High Frequency components can be incorporated within the same framework such that the Low Frequency (or trend) is well protected from the High Frequency(or coloured noise component) especially in the long run.

1.3 Organisation of the Thesis

In this dissertation, chapter one gives a historical review of the basic techniques involved in the development of the G.E.W.R. Methodology. For a better understanding of the developed theory, definitions of the terminology to be used are given.

In the second chapter Exponentially Weighted Regression (E.W.R) theory is presented. Some basic results together their limiting forms are given. Simple and Higher Order smoothing techniques are discussed.

In chapter three, a Linear Dynamic System for the analysis of time series with a white noise component is considered. For Dynamic Linear Models (D.L.M) setting procedures for the system matrices are given. For unknown noise variances, an on line Bayesian Variance Learning Procedure is given. The relationship between E.W.R. and D.L.M. is explained. E.W.R. type D.L.M's are given.

In chapter four, Autoregressive and Moving Average (ARMA) type noise processes are considered. The recurrence relations for the covariance and precision matrices are given.

In chapter five, the main theory of G.E.W.R. is presented through various theorems and their corollaries. The Fundamental and the Characterisation theorems of G.E.W.R. are more elaborately stated and proved. The recurrence relations for updating the dynamic systems are given, together their Forecast Functions. The recurrence relations given do not generally require matrix inversion.

In the sixth chapter, a dynamic model representation of G.E.W.R. is considered. Some theorems, corollaries and examples are given. Dynamic Linear Models and Normal Discount Bayesian Models of G.E.W.R. type are also discussed. A procedure to construct State Space type models is presented. For theory and practice the recurrence relations for the transformation matrices that transform a dynamic system of the canonical form to a diagonal form and vice-versa are derived. For an unknown noise variance, an on line Bayesian Learning Procedure for variance learning is developed.

In chapter seven, some practical aspects of the G.E.W.R.

Methodology are considered. A new identification procedure to identify appropriate G.E.W.R. models is given. In order to demonstrate the theory in practice, the applications of G.E.W.R. to three simulated data sets and four real life economic and industrial data sets consisting of time series driven by coloured noise processes are presented.

In chapter eight, a brief summary, followed by closing remarks and suggestions for further research are given.

Finally, relevant references are listed in the alphabetical order.

1.4 Preliminaries

1.4.1 Stochastic Process

A stochastic process is a mathematical abstraction of an empirical process generated by probabilistic laws. More precisely, it is a family of random variables $\{X_t\}_{t\varepsilon T}$ defined over some range or time space T.

The process is classified as a Discrete Stochastic Process if the time space T is discrete, i.e. T is defined as a space of an infinite or finite sequence

For example:

$$T = \{0, \pm 1, \pm 2, \ldots\ldots\}$$

or

$$T = \{0, 1, 2, \ldots\ldots\ldots\}$$

and classified as a Continuous Stochastic Process if T is continuous, i.e. T is defined over some interval

For example:

$$T = \{ t: \quad -\infty < t < \infty \}$$

or

$$T = \{ t: \quad 0 \leq t < \infty \}$$

Further the process $\{X_t\}_{t\,\varepsilon\,T}$ is Stationary if the probability structure or Statistics are time invariant; otherwise it is called Non-Stationary.

1.4.2 Markov Process

A stochastic process $\{X_t\}_{t \in T}$ is called a Markov Process for any set of n time points $t_1, t_2, \ldots, t_n$, defined on the time space T, if the conditional distribution of X_{tn} given the values of $X_{t1}, X_{t2}, \ldots, X_{t(n-1)}$ depends only on $X_{t(n-1)}$, the immediately preceding value.

More precisely for any real numbers $X_1, X_2, \ldots, X_n$

$$P\{X_{tn} \leq X_n \mid X_{t1} = X_1, X_{t2} = X_2, \ldots, X_{t(n-1)} = X_{n-1}\}$$

$$= P\{X_{tn} \leq X_n \mid X_{t(n-1)} = X_{n-1}\}.$$

In non mathematical language we can say that:

given the present condition of the process, the future is independent of the past.

Further x is said to be a State of a stochastic process $\{X_t\}_{t \in T}$ if there exists a time point t in the time space T such that:

$$P\{x - \# < X_t < x + \#\} > 0 \quad ; \text{for all } \# > 0.$$

The set of such possible States constitute the State Space of the process.

1.4.3 Noise

A stochastic process $\{\delta_t\}_{t \in T}$ generated by some random disturbances is called simply noise or a noise process.

Often it is classified as white noise or a coloured noise.

The term white noise is applied to the case where the members of $\{\delta_t\}_{t \in T}$ are mutually uncorrelated with zero mean and common variance. Statistically it is stationary having constant spectral density at all frequencies and zero autocorrelation except at t=0.

For empirical studies, if required, the covariance function of white noise can be symbolically represented by a Dirac Delta or Impulse Function $\Delta(t)$, where

$$\int_{-\infty}^{\infty} \Delta(t)\, dt = 1$$

and for any function $\gamma(t)$ continuous at $t = 0$

$$\int_{-\infty}^{\infty} \gamma(t) \; \Delta(t)\, dt = \gamma(0).$$

The noise which is not white, for brevity is called a coloured noise.

The exact form of the noise is quite difficult to know. In this dissertation, it is assumed that the white noise is normally distributed with mean zero and some variance σ_{δ}^{2} and coloured noise follow ARMA(p,q) process.

1.4.4 Parsimony

The term Parsimony refers to the principles:

1) of using the smallest possible number of parameters for an adequate representation of a stochastic process. This is called Parametric Parsimony. The order of the Parsimony depends upon number of free parameters to be estimated from the data.

2) of arguing that for a fairly general class of accuracy criteria, a simpler structure of the dynamics of a model is asymptotically better, for a Statistically efficient estimation method. This is called Conceptual Parsimony.

These two principles which are part and parcel of model building, play quite a vital role, especially in commercial modelling and forecasting.

For more discussion see Roberts-Harrison (1982) and Stoica-Soderstrom (1982).

2

EXPONENTIALLY WEIGHTED REGRESSION

E.W.R.

2.1 Introduction

Exponentially Weighted Regression (E.W.R.) is an exponential smoothing technique based on the 'Fading Memory Logic', originally advocated by Holt (1957). This states that relatively more importance should be given to information received from more recent data than from earlier data.
Essentially the basic idea is that the importance of an observation with respect to current estimation, decays exponentially with its age. In the light of this logic the weights assigned to the past time series realizations decay monotonically over time, thus giving maximum weight to the information received from the most recent data. The weights so allocated act as a low pass filter.

2.2 Simple Exponential Smoothing

The simplest form of exponential smoothing is described below:

At time t-i (i=0,..., t-1), let a discrete stochastic phenomenon be locally modelled as

$$Y_{t-i} = \theta + \delta_{t-i} \qquad (2.2.0.1)$$

where Y_{t-i} is an observation, θ is a stochastic parameter to be estimated and $\{\delta_{t-i}\}$ is a Gaussian White Noise sequence with mean zero and variance σ^2_δ .

Given a memory or discount factor $0 < \beta < 1$ and data $D_t = y_t, y_{t-1}, \ldots, y_1$, the parameter θ ,at time t, is estimated by m_t, where m_t is that value of θ which minimizes the discounted sum of squares

$$S = \sum_{i=0}^{t-1} \beta^i (y_{t-i} - \theta)^2 . \qquad (2.2.0.2)$$

Minimizing S with respect to θ we get

$$\frac{dS}{d\theta}\bigg|_{\theta=m_t} = 0 = -2 \sum_{i=0}^{t-1} \beta^i (y_{t-i} - \theta) \bigg|_{\theta=m_t}$$

or

$$m_t \sum_{i=0}^{t-1} \beta^i = \sum_{i=0}^{t-1} \beta^i y_{t-i} = \sum_{i=0}^{t-1} (\beta B)^i y_t$$

where B is the backward shift operator such that $By_t = y_{t-1}$ and $Bm_t = m_{t-1}$. Using the standard result for the sum of Geometric series we can write the above expression as

$$m_t(1-\beta^t)/(1-\beta) = [\{1-(\beta B)^t\} /(1-\beta B)] y_t.$$

Re-arranging the terms we get

$$(1-\beta B)m_t = (1-\beta)[\{ 1-(\beta B)^t\} /(1-\beta^t)] y_t.$$

In the limit $t\to\infty$, $\beta^t \to 0$ and $(\beta B)^t \to 0$, so in the limit we get

$$(1-\beta B)m_t = (1-\beta)y_t \tag{2.2.0.3}$$

or

$$m_t = \beta m_{t-1} + (1-\beta)y_t. \tag{2.2.0.4}$$

This result can be written in various other forms. For example, since $0 < \beta < 1$ and $(1-\beta B)$ is invertible, we can write (2.2.0.3) as

$$m_t = (1-\beta)(1-\beta B)^{-1}y_t = \alpha \sum_{i=0}^{\infty} \beta^i y_{t-i} \tag{2.2.0.5}$$

where $\alpha = 1-\beta$.

Defining the k-steps ahead forecast function for the model (2.2.0.1) by

$$F_t(k) = m_t \quad ; k = 1,2,... \tag{2.2.0.6}$$

and defining the one step ahead forecast errors by $e_t = y_t - m_{t-1}$ we can write (2.2.0.4) as

$$m_t = m_{t-1} + \alpha e_t \tag{2.2.0.7}$$

or $\quad (1-B)m_t = \alpha e_t$.

Substituting the value of m_t from (2.2.0.5) into (2.2.0.7) and simplifying we get

$$(1-B)(1-\beta B)^{-1}\, y_t = e_t$$

or $$(1-B)y_t = (1-\beta B)e_t . \qquad (2.2.0.8)$$

This shows that in the limit $t\to\infty$ the simple exponential smoothing model(2.2.0.1) is equivalent to the first order model of Brown(1962) and the subsets of Box-Jenkins(1976) ARIMA(0,1,1) model and the first order model of Holt(1957). The two latter models would allow values of β such that $|\beta|<1$. Muth(1960) has shown that such a model yields optimum one step ahead forecasts in the minimum mean square sense (M.M.S.E.).

Simple exponential smoothing is attractive, parsimonious and widely applied due to low computational costs. However, it is best suited for time series having essentially a horizontal pattern. When dealing with a trend or even a seasonal pattern, it becomes necessary to deal with higher order smoothing. This concept is introduced in the following sections.

2.3 Theorem (T1)

For a sequence Y_t assuming at time t-i, a local model

$$Y_{t-i} = \underline{f}_{t-i}\, \underline{\theta} + \delta_{t-i} \quad ; \text{ for } i=0,1,\ldots,t-1$$

where Y_{t-i}, $\underline{\theta}$ and δ_{t-i} are as defined in section (2.2); and $\underline{f}_{t-i}$ are (1xn) row vectors of some known functions of independent variables. For detail see Appendix K.

Considering:

i) the k-steps ahead forecast function

$$F_t(k) = E(\, Y_{t+k} |\, D_t\,) = \underline{f}_{t+k}\, \underline{m}_t$$

for a given data $D_t = y_t,\ldots,y_1$ and $\underline{m}_t$ as an estimate of $\underline{\theta}$ at time time t;

ii) one step ahead forecast error

$$e_t = y_t - \underline{f}_t\, \underline{m}_{t-1};$$

and defining

$$\underline{Q}_t = \sum_{i=0}^{t-1} \beta^i \underline{f}_{t-i}' \underline{f}_{t-i}$$

a (nxn) system matrix of full rank, consisting of discounted sums of squares and cross products of the independent variables computed at time t, where the discount factor β is such that $0 < \beta < 1$; then $\underline{m}_t$ and $\underline{Q}_t$ may be evaluated recursively through the recurrence relations

$$\underline{m}_t = \underline{m}_{t-1} + \underline{Q}_t^{-1} \underline{f}_t' e_t$$

$$\underline{Q}_t = \underline{f}_t' \underline{f}_t + \beta \underline{Q}_{t-1} .$$

Proof :

Let at time t, $\underline{m}_t$ be that value of $\underline{\theta}$ which minimizes the discounted sum of squares

$$S = \sum_{i=0}^{t-1} \beta^i (y_{t-i} - \underline{f}_{t-i}\underline{\theta})^2. \qquad (2.3.0.1)$$

Minimizing with respect to $\underline{\theta}$, we get

$$\left.\frac{dS}{d\underline{\theta}}\right|_{\underline{m}_t} = \underline{0} = -2 \sum_{i=0}^{t-1} \beta^i \underline{f}_{t-i}' (y_{t-i} - \underline{f}_{t-i}\underline{\theta}) \Big|_{\underline{m}_t}$$

or

$$\sum_{i=0}^{t-1} \beta^i \underline{f}_{t-i}' \underline{f}_{t-i} \underline{m}_t = \sum_{i=0}^{t-1} \beta^i \underline{f}_{t-i}' y_{t-i} \qquad (2.3.0.2)$$

$$= \underline{H}_t \text{ say.}$$

Using the above definition of $\underline{Q}_t$, we get

$$\underline{Q}_t \underline{m}_t = \underline{H}_t . \qquad (2.3.0.3)$$

Now

$$\underline{H}_t = \sum_{i=0}^{t-1} \beta^i \underline{f}_{t-i}' y_{t-i} = \underline{f}_t' y_t + \beta \sum_{i=1}^{t-1} \beta^{i-1} \underline{f}_{t-1}' y_{t-i}$$

Shifting the origin we get

$$\underline{H}_t = \underline{f}_t^{'} y_t + \beta \sum_{i=0}^{t-2} \beta^i \underline{f}_{t-1-i}^{'} y_{t-1-i}$$

$$= \underline{f}_t^{'} y_t + \beta \underline{H}_{t-1} \qquad (2.3.0.4)$$

Similarly

$$\underline{Q}_t = \underline{f}_t^{'} \underline{f}_t + \sum_{i=0}^{t-2} \beta^i \underline{f}_{t-1-i}^{'} \underline{f}_{t-1-i}$$

$$= \underline{f}_t^{'} \underline{f}_t + \beta \underline{Q}_{t-1} . \qquad (2.3.0.5)$$

Now using backward shift operator B on (2.3.0.3) we get

$$B(\underline{Q}_t \underline{m}_t) = B\underline{H}_t$$

or $$\underline{Q}_{t-1} \underline{m}_{t-1} = \underline{H}_{t-1} . \qquad (2.3.0.6)$$

Comparing (2.3.0.6) with (2.3.0.4) we get

$$\underline{H}_t = \underline{f}_t^{'} y_t + \beta \underline{Q}_{t-1} \underline{m}_{t-1} .$$

Inserting this value of $\underline{H}_t$ in (2.3.0.3) we get

$$\underline{Q}_t \underline{m}_t = \underline{f}_t^{'} y_t + \beta \underline{Q}_{t-1} \underline{m}_{t-1} .$$

Now from the definition of e_t we know that

$$y_t = e_t + \underline{f}_t \underline{m}_{t-1} .$$

So

$$\underline{Q}_t \underline{m}_t = \underline{f}_t^{'} (\underline{f}_t \underline{m}_{t-1} + e_t) + \beta \underline{Q}_{t-1} \underline{m}_{t-1}$$

$$= (\underline{f}_t^{'} \underline{f}_t + \beta \underline{Q}_{t-1}) \underline{m}_{t-1} + \underline{f}_t^{'} e_t$$

$$= \underline{Q}_t \underline{m}_{t-1} + \underline{f}_t^{'} e_t \quad ;$$

or

$$\underline{m}_t = \underline{m}_{t-1} + \underline{Q}_t^{-1} \underline{f}_t^{'} e_t . \qquad (2.3.0.7)$$

■

2.3.1 Corollary 1

For a time series, it is usual to adopt a moving parameterisation through a matrix $\underline{G}$ such that at time t, for an integer k

$$Y_{t+k} = \underline{f}_k \underline{\theta} + \delta_{t+k} \qquad (2.3.1.1)$$

where $\underline{f}_k = \underline{f}_0 \underline{G}^k$, $(\underline{f}_0', \underline{f}_1', \ldots, \underline{f}_{n-1}')$ is of rank n and $\underline{G}$ is a (nxn) time shift or transition matrix of full rank with non zero elements on its main diagonal. (see Appendix K). The eigenvalues of $\underline{G}$ determine the form of the forecast function so that the k - steps ahead forecast function is

$$F_t(k) = E(Y_{t+k} | D_t) = \underline{f}_k \underline{m}_t$$

$$= \underline{f}_0 \underline{G}^k \underline{m}_t. \qquad (2.3.1.2)$$

These definitions modify the expressions (2.3.0.4), (2.3.0.5) and (2.3.0.7) for $\underline{H}_t$, $\underline{Q}_t$ and $\underline{m}_t$ respectively to

$$\underline{H}_t = \underline{f}_0' y_t + \beta(\underline{G}')^{-1} \underline{H}_{t-1} \qquad (2.3.1.3)$$

$$\underline{Q}_t = \underline{f}_0' \underline{f}_0 + \beta(\underline{G}')^{-1} \underline{Q}_{t-1} \underline{G}^{-1} \qquad (2.3.1.4)$$

$$\underline{m}_t = \underline{G}\, \underline{m}_{t-1} + \underline{Q}_t^{-1} \underline{f}_0' e_t \qquad (2.3.1.5)$$

where $e_t = y_t - \underline{f}_0 \underline{G}\, \underline{m}_{t-1}$.

When $\underline{G} = \underline{I}$, these results obviously reduce to (2.3.0.4) (2.3.0.5) and (2.3.0.7) respectively if $\underline{f}_t = \underline{f}_0$.

2.3.2 Corollary 2

Writing $\underline{f} = \underline{f}_0$ and defining $\underline{Q}_t^{-1} \underline{f}' = \underline{A}_t$ the result (2.3.1.5) can be written as:

$$\underline{m}_t = \underline{G}\,\underline{m}_{t-1} + \underline{A}_t\, e_t . \qquad (2.3.2.1)$$

2.3.3 Limiting Results

For a discount factor β such that $0 < \beta < \min_i |\lambda_i^2|$ where λ_i are the eigenvalues of $\underline{G}$, the following limits exist.

$$\lim_{t\to\infty} \underline{Q}_t = \underline{Q} = \underline{f}'\underline{f} + (\underline{G}')^{-1}\,\underline{Q}\,\underline{G}^{-1} \qquad (2.3.3.1)$$

$$\lim_{t\to\infty} \underline{A}_t = \underline{A} = \underline{Q}^{-1}\,\underline{f}' . \qquad (2.3.3.2)$$

The limit (2.3.3.1) is a special case of (5.2.3.3.1) where the current value of $\underline{f}_t$ considered is $\underline{f}$. The limit (2.3.3.2) obviously follows. Analytically, if we wish, the limiting values of $\underline{Q}_t$ and $\underline{A}_t$ can be easily evaluated by using above expressions.

Example

For a second order (n = 2) form of a local model defined earlier, let in a time series case

$$\underline{f} = (1\;\;0)\;,\quad \underline{G} = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{bmatrix} \quad \text{and} \quad \underline{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} .$$

The limiting value of $\underline{Q}$ is evaluated by quoting the respective values of $\underline{f}$, $\underline{G}$, $\underline{G}^{-1}$ and $\underline{Q}$ in (2.3.3.1). Thus

$$\begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}(1\;\;0) + \beta \begin{bmatrix} \lambda_1^{-1} & 0 \\ -\lambda_1^{-1}\lambda_2^{-1} & \lambda_2^{-1} \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} \lambda_1^{-1} & -\lambda_1^{-1}\lambda_2^{-1} \\ 0 & \lambda_2^{-1} \end{bmatrix}$$

(2.3.3.3)

Solving in terms of the q's we get

$$q_{11} = \lambda_1^2 \;/\; (\lambda_1^2 - \beta)$$

$$q_{12} = -\lambda_1 \beta \;/ (\lambda_1\lambda_2 - \beta)(\lambda_1^2 - \beta)$$

$$q_{22} = \beta(\lambda_1\lambda_2 + \beta) \;/ (\lambda_1^2 - \beta)(\lambda_2^2 - \beta)(\lambda_1\lambda_2 - \beta)$$

hence

$$\underline{Q} = (\lambda_1^2 - \beta)^{-1} \begin{bmatrix} \lambda_1^2 & , -\lambda_1\beta \;/ (\lambda_1\lambda_2 - \beta) \\ -\lambda_1\beta \;/ (\lambda_1\lambda_2 - \beta), & (\lambda_1\lambda_2 + \beta)\beta / [(\lambda_2^2 - \beta)(\lambda_1\lambda_2 - \beta)] \end{bmatrix} ;$$

(2.3.3.4)

the inverse of which is

$$\underline{Q}^{-1} = \{(\lambda_1\lambda_2 - \beta)/\lambda_2^2\} \begin{bmatrix} (\lambda_1\lambda_2 + \beta)/\lambda_1^2 & (\lambda_2^2 - \beta)/\lambda_1 \\ (\lambda_2^2 - \beta)/\lambda_1 & (\lambda_1\lambda_2 - \beta)(\lambda_2^2 - \beta)/\beta \end{bmatrix} .$$

(2.3.3.5)

Using the result of $\underline{Q}^{-1}$ in (2.3.3.2) we get the limiting value of the updating vector $\underline{A}_t$, i.e.

$$\underline{A} = \underline{Q}^{-1} \underline{f} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \qquad (2.3.3.6)$$

with

$$A_1 = \{ (\lambda_1\lambda_2)^2 - \beta^2 \} / (\lambda_1\lambda_2)^2$$

and

$$A_2 = (\lambda_1\lambda_2 - \beta)(\lambda_2^2 - \beta)/\lambda_1\lambda_2^2 .$$

Special Cases

i) If $\lambda_1 = \lambda_2 = \lambda$,i.e. $\underline{G} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$ we get the limiting results as

$$\underline{Q} = (\lambda^2 - \beta)^{-1} \begin{bmatrix} \lambda^2 & -\lambda\beta / (\lambda^2 - \beta) \\ -\lambda\beta / (\lambda^2 - \beta) & (\lambda^2 + \beta)\beta / (\lambda^2 - \beta)^2 \end{bmatrix} . \qquad (2.3.3.7)$$

$$\underline{Q}^{-1} = (\lambda^2 - \beta)/\lambda^2 \begin{bmatrix} (\lambda^2 + \beta)/\lambda^2 & (\lambda^2 - \beta)/\lambda \\ (\lambda^2 - \beta)/\lambda & (\lambda^2 - \beta)^2 / \beta \end{bmatrix} \qquad (2.3.3.8)$$

and

$$\underline{A} = \underline{Q}^{-1} \, \underline{f} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$$

with the values of A_1 and A_2 as

$$A_1 = (\lambda^4 - \beta^2)/\lambda^4$$

and

$$A_2 = (\lambda^2 - \beta)^2 / \lambda^3 . \qquad (2.3.3.9)$$

ii) If $\lambda_1 = \lambda_2 = 1$ i.e. $G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ the Linear Growth case, then

$$\underline{Q} = (1 - \beta)^{-1} \begin{bmatrix} 1 & -\beta/(1-\beta) \\ -\beta/(1-\beta) & (1+\beta)\beta/(1-\beta)^2 \end{bmatrix} , \qquad (2.3.3.10)$$

$$\underline{Q}^{-1} = (1-\beta)\begin{bmatrix} 1+\beta & 1-\beta \\ 1-\beta & (1-\beta)^2/\beta \end{bmatrix} \qquad (2.3.3.11)$$

and $\quad \underline{A} = \underline{Q}^{-1}\underline{f} = (\, A_1 \quad A_2 \,)'$ where

$$A_1 = (1-\beta^2) \quad \text{and} \quad A_2 = (1-\beta)^2 \qquad (2.3.3.12)$$

iii) If $\lambda_1 = 1$, $\lambda_2 = \lambda$; i.e. $\underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & \lambda \end{bmatrix}$ then we get

$$\underline{Q} = (1-\beta)^{-1}\begin{bmatrix} 1 & -\beta/(\lambda-\beta) \\ -\beta/(\lambda-\beta) & \beta(\lambda+\beta)/(\lambda^2-\beta)(\lambda-\beta) \end{bmatrix}$$

(2.3.3.13)

$$\underline{Q}^{-1} = (\lambda-\beta)/\lambda \begin{bmatrix} \lambda+\beta & \lambda^2-\beta \\ \lambda^2-\beta & (\lambda-\beta)(\lambda^2-\beta)/\beta \end{bmatrix}$$

(2.3.3.14)

and $\quad \underline{A} = \underline{Q}^{-1}\underline{f} = (\, A_1 \quad A_2 \,)'$ where

$$A_1 = (\lambda^2-\beta^2)/\lambda^2 \quad \text{and} \quad A_2 = (\lambda-\beta)(\lambda^2-\beta)/\lambda^2$$

■ (2.3.3.15)

A procedure for finding the limiting values of $\underline{Q}_t$ and $\underline{A}_t$ is discussed above for second order E.W.R. models in their canonical forms. For the diagonal cases or any other form of system and for higher order cases, the limiting results of $\underline{Q}_t$ and $\underline{A}_t$ can be evaluated following the same procedure.

The expression (2.3.1.5) in its limiting form can be written as

$$\underline{m}_t = \underline{G}\ \underline{m}_{t-1} + \underline{Q}^{-1}\underline{f}'\, e_t = \underline{G}\ \underline{m}_{t-1} + \underline{A}\ e_t\ . \qquad (2.3.3.16)$$

2.4 Brown's Multiple Exponential Smoothing

The multiple or higher order exponential smoothing operation is carried out repeatedly on a time series assuming higher degree polynomial models. Brown (1962) suggested an nth order smoothing operation as

$$m_t^{(n)} = \alpha m_t^{(n-1)} + \beta m_{t-1}^{(n)} \qquad (2.4.1)$$

where the superscripts represent the order of smoothing operation, $0 < \beta < 1$, $\alpha = 1 - \beta$, $m_t^{(1)} = mt$, $m_t^{(0)} = y_t$ and $m_1^{(0)} = y_1$.

The basic idea behind such a smoothing operation is to apply single or simple exponential smoothing (2.2.0.4) first to the original data and then repeat the smoothing operation on the smoothed series obtained from the first smoothing. We go on repeating the smoothing operation on the smoothed series of the lower order until we get the smoothed values of the desired order. The smoothed values are then used to estimate the elements $(\theta_1, \ldots, \theta_n)$ of the unknown stochastic parameter vector $\underline{\theta}$ by $\underline{m}_t = (m_1, \ldots, m_n)'_t$ following the procedure described by Brown (1962). However, an improved estimate of the level or intercept at the current origin $m_{1,t}$ is evaluated by linearly combining the lower and higher order smoothed values as

$$m_{1,t} = \sum_{j=1}^{n} (-1)^{j-1} \binom{n}{j} m_t^{(j)}\ . \qquad (2.4.2)$$

The one step ahead limiting forecasts obtained by the method described are optimum in the minimum mean square error (M.M.S.E.) sense and are equivalent to those derived from a corresponding subset of Box-Jenkins(1976) ARIMA(0,n,n) model, i.e.

$$(1 - B)^n y_t = (1 - \beta B)^n e_t\ . \qquad (2.4.3)$$

The general ARIMA(0,n,n) model is

$$(1 - B)^n y_t = \prod_{i=1}^{n} (1 - \eta_i B) e_t \qquad (2.4.4)$$

with the restriction that $|\eta_i| < 1$ for $i=1,...,n$, whereas the above subset restricts $0< \eta_i = \beta <1$ for all i. For related discussion see Godolphin-Harrison (1975).

There are certain disadvantages of Brown's multiple exponential smoothing procedure which are:

i) The procedure for estimating the elements of the parameter vector $\underline{\theta}_t$ is quite cumbersome and computationally inefficient as the estimates of the parameters need to be updated at the end of each period with the arrival of new observation and it is not easy to compute by the Brown's procedure.

ii) The limiting result is useful only if enough data is available and the process has not fundamentally changed.

These problems may be overcomed if instead of using indirect smoothing described by Brown, direct smoothing based on exponentially weighted least square criterion is used, such as described by E.W.R. theorem (T1)(2.3) and the Dynamic Exponentially weighted Regression (D.E.W.R.) models introduced in chapter 3, where recurrence relations for the updating vector $\underline{A}_t$ and inversion of the covariance matrix $\underline{Q}_t$ are developed.

2.5 The Discount Factor (β)

The memory or discount factor β defines the rate at which the precision of an estimate of a parameter is lost. It plays quite a significant role in time series analysis and forecasting. It helps the models to track the parameter variations in the time series by discounting the past. Its value, lying between zero and one is chosen keeping in mind certain features associated with values of β near 0 and 1. A low value of β discounts the past data very fast, whereas, a high value of β discounts it very slowly. The consequencs are:

If the past data is discounted:

i) Too Fast, i.e. the value of β selected is too low, the system will over respond to the random noise, introduce negative correlation to the one step ahead forecast errors and increase the forecast variance.

ii) Too Slowly, i.e. the value of β selected is too close to one, the system will respond poorly to systematic process changes, introduce positive correlation to the one step ahead forecast errors and decrease the forecast variance (see Harrison-Scott(1965)). The estimates of parameters in a nearly constant model will be good but the estimator may fail to respond adequately to the parameter variations required by models which only locally capture the process description.

The two stated approaches highlights the dangers of selecting the extreme value of the discount factor β. A reasonable value of β that helps us to obtain forecasts with uncorrelated one step ahead forecast errors is recommended.

The optimum selection of discount factor β may depend upon some other factors, such as:

a) Nature of data, e.g.

i) seasonal
ii) non-seasonal.

b) Type of noise present in time series, e.g.

i) white noise
ii) coloured noise.

c) Type of model being used, e.g.

i) E.W.R. type model
ii) G.E.W.R. type model.

d) Model scheme being used, e.g.

i) independent models for short term and long term forecasts
ii) joint modelling scheme that incorporates both short and long term forecasts.

In time series where usually moving parameterisation is adopted through a (nxn) transition matrix $\underline{G}$, an additional restriction on the discount factor is imposed such that

$$0 < \beta < \min | \lambda_i^2 | \qquad \text{for } i=1,\ldots, n \qquad (2.5.1)$$

where $\lambda_1,\ldots, \lambda_n$ are non zero eigenvalues of the transition matrix $\underline{G}$, which are not necessarily distinct. This restriction ensures the convergence of the parameters to a limiting form and during the process of convergence the order of the model remains same. This situation is discussed through the G.E.W.R. theorem (T3) (5.3.1) given in chapter five.

In cases where the restriction (2.5.1) is not fully met so that

$$0 < \beta < \min | \lambda_i^2 | \qquad \text{for } i=1,\ldots, m \qquad (2.5.2)$$

$$\beta > | \lambda_i^2 | \qquad \text{for } i=m+1,\ldots, n \qquad (2.5.3)$$

the order of the model reduces from n to m during the process of the convergence of the parameters. This result is derived in theorem (T4) (5.3.5).

There are many ways to select the discount factor in the light of the various aspects and the constraints mentioned earlier. One way of selecting β approximately is to think about it in terms of half times. If the information content of a point is decreased to half its initial value after τ periods then for $\tau \geq 1$

$$\beta \cong (0.5)^{2/T} = (3\tau - 1)/ (3\tau + 1)$$

where T is the length of the time series available for analysis. For more discussion see Harrison-Johnston (1983).

<u>Comment</u>

The restriction (2.5.1) apply only in the time series case as in this case moving parameterisation is considered through the transition matrix $\underline{G}$, having n λ_i eigenvalues. It does not apply to ordinary case as $\underline{G} = \underline{I}$ in this case.

3

LINEAR DYNAMIC SYSTEMS

3.1 Introduction

In the previous chapter I discussed some basic work, concerned with Exponential Weighted Regression (E.W.R.). This simple but useful concept is developed further here. In section (2.3) we have seen that, in order to update $\underline{m}_t$, an estimate of parameter vector $\underline{\theta}$ at time t, through the relation (2.3.0.7) or relation (2.3.1.5), every time, we need to invert the matrix $\underline{Q}_t$. This is quite cumbersome. To overcome this problem, here, dynamic recurrence relations are developed which help us to evaluate $\underline{Q}_t^{-1}$ recursively. First, an inversion identity is presented in section 2 for matrices and vector-matrix equations.

In section 3 a dynamic form of E.W.R., both for the ordinary and time series cases is given.

In section 4 a wide class of Dynamic Linear Models, introduced by Harrison - Stevens (1976) is presented. Some of the most commonly used special cases are discussed with new developments.

In section 5 E.W.R. type Dynamic Linear Models are given.

In section 6 Seasonal Models are discussed.

3.2 Inversion Identity

3.2.1 General Matrix Identity

A well known inversion identity (see Lindley-Smith(1972)) which is useful in dealing with Linear Dynamic System is given below.

Let $\underline{A}$, $\underline{B}$, $\underline{C}$ and $\underline{D}$ be matrices of appropriate dimensions of which atleast $\underline{A}$, $\underline{C}$ and $\underline{D}$ are invertible, then

$$(\underline{A}^{-1} + \underline{B}'\underline{C}\,\underline{B})^{-1} = \underline{A} - \underline{A}\,\underline{B}'(\underline{C}^{-1} + \underline{B}\,\underline{A}\,\underline{B}')^{-1} \underline{B}\,\underline{A}$$

$$= (\underline{I} - \underline{A}\,\underline{B}'\underline{D}^{-1}\,\underline{B})\,\underline{A} \quad , \qquad (3.2.1.1)$$

where $\underline{D} = (\underline{C}^{-1} + \underline{B}\ \underline{A}\ \underline{B}')$. (3.2.1.2)

3.2.2 Vector-Matrix Mixed Identity

As a special case of (3.2.1), if $\underline{B}$ is a row vector, written as $\underline{b}$; c and d are scalars instead of matrices $\underline{C}$ and $\underline{D}$ respectively, then

$$(\underline{A}^{-1} + c\ \underline{b}'\underline{b})^{-1} = (\underline{I} - \underline{A}\ \underline{b}'\underline{b}\ d^{-1})\ \underline{A} \qquad (3.2.2.1)$$

where $d = (c^{-1} + \underline{b}\ \underline{A}\ \underline{b}')$. (3.2.2.2)

3.3 Dynamic E.W.R. (D.E.W.R.)

3.3.1 Ordinary Form of D.E.W.R.

In order to find the estimate $\underline{m}_t$ of the parameter vector $\underline{\theta}$ by the recurrence relation (2.3.0.7), i.e.

$$\underline{m}_t = \underline{m}_{t-1} + \underline{Q}_t^{-1}\ \underline{f}_t'\ e_t \qquad (3.3.1.1)$$

we need to invert the covariance matrix $\underline{Q}_t$ at every time point t. This is quite cumbersome. In order to overcome this problem dynamic recurrence relations for $\underline{Q}_t^{-1}$ and the updating vector $\underline{A}_t$ are developed. These recurrence relations form a part of a dynamic system where no matrix inversion is required. The matrix $\underline{Q}_t^{-1}$ is updated automatically in a recursive manner, once a priori information is provided to the dynamic system.

For a local model (2.3), i.e.

$$Y_{t-i} = \underline{f}_{t-i}\ \underline{\theta} + \delta_{t-i} \quad ; i = 0,\dots,t-1 \qquad (3.3.1.2)$$

where at time t, the components of the model, i.e. Y_{t-i} , $\underline{f}_{t-i}$, $\underline{\theta}$ and δ_{t-i} are as defined in section (2.3).

Following the definition of the k-steps ahead forecast function (2.3) , i.e.

$$F_t(k) = E(Y_{t+k} \mid D_t) = \underline{f}_{t+k}\, \underline{m}_t \qquad (3.3.1.3)$$

and one step ahead forecast error

$$e_t = y_t - \underline{f}_t\, \underline{m}_{t-1} \qquad (3.3.1.4)$$

the vector $\underline{m}_t$ is recursively evaluated as

$$\underline{m}_t = \underline{m}_{t-1} + \underline{A}_t\, e_t \qquad (3.3.1.5)$$

where the updating vector $\underline{A}_t$ is computed through the recurrence relations

$$\underline{K}_t = \underline{Q}_{t-1}^{-1} / \beta \qquad (3.3.1.6)$$

$$\underline{A}_t = \underline{K}_t\, \underline{f}_t' \,(1 + \underline{f}_t\, \underline{K}_t\, \underline{f}_t')^{-1} \qquad (3.3.1.7)$$

$$\underline{Q}_t^{-1} = (\underline{I} - \underline{A}_t\, \underline{f}_t)\, \underline{K}_t \qquad (3.3.1.8)$$

Proof:

From equation (2.3.0.5) we know that

$$\underline{Q}_t = \beta \underline{Q}_{t-1} + \underline{f}_t'\, \underline{f}_t \qquad (3.3.1.9)$$

and from (2.3.2) for $\underline{f}_t = \underline{f}$

$$\underline{A}_t = \underline{Q}_t^{-1}\, \underline{f}_t' \,. \qquad (3.3.1.10)$$

We also know from (3.3.1.6) that

$$\underline{K}_t^{-1} = \beta \, \underline{Q}_{t-1} \, . \tag{3.3.1.11}$$

Substituting (3.3.1.11) into (3.3.1.9) we get

$$\underline{Q}_t = \underline{K}_t^{-1} + \underline{f}_t' \, \underline{f}_t \, . \tag{3.3.1.12}$$

Using the matrix inversion identity (3.2.2.1) we can write

$$\underline{Q}_t^{-1} = \underline{K}_t \, \underline{f}_t' \, (1 + \underline{f}_t \, \underline{K}_t \, \underline{f}_t')^{-1} \, \underline{f}_t \, \underline{K}_t \, .$$

Writing $\hat{Y}_t = 1 + \underline{f}_t \, \underline{K}_t \, \underline{f}_t'$ (3.3.1.13)

and re-arranging the terms we have

$$\underline{Q}_t^{-1} = \{ \underline{I} - \underline{K}_t \, \underline{f}_t' \, (\hat{Y}_t)^{-1} \, \underline{f}_t \} \, \underline{K}_t \tag{3.3.1.14}$$

Post multiplying (3.3.1.10) by $\underline{f}_t$ we get

$$\underline{A}_t \, \underline{f}_t = \underline{Q}_t^{-1} \, \underline{f}_t' \, \underline{f}_t .$$

Substituting here the value of $\underline{f}_t' \, \underline{f}_t$ from (3.3.1.12) we get

$$\underline{A}_t \, \underline{f}_t = \underline{Q}_t^{-1} \, (\underline{Q}_t - \underline{K}_t^{-1})$$

$$= \underline{I} - \underline{Q}_t^{-1} \, \underline{K}_t^{-1}$$

Re-arranging the terms we get

$$\underline{Q}_t^{-1}\ \underline{K}_t^{-1} = \underline{I} - \underline{A}_t\ \underline{f}_t\ .$$

Post multiplying both sides by $\underline{K}_t$ we get

$$\underline{Q}_t^{-1} = (\ \underline{I} - \underline{A}_t\ \underline{f}_t\ )\ \underline{K}_t\ . \qquad (3.3.1.15)$$

Comparing this expression with (3.3.1.14) we get

$$(\ \underline{I} - \underline{A}_t\ \underline{f}_t\ )\ \underline{K}_t = \{\ \underline{I} - \underline{K}_t\ \underline{f}_t'\ (\ \hat{Y}_t\ )^{-1}\ \underline{f}_t\}\ \underline{K}_t\ .$$

Simplifying this expression we get

$$\underline{A}_t = \underline{K}_t\ \underline{f}_t'\ (\ \hat{Y}_t\ )^{-1}\ . \qquad (3.3.1.16)$$

This completes the proof.

Re-writing the recurrence relations in order we have

$$\underline{K}_t = \underline{Q}_{t-1}^{-1}\ /\beta$$

$$\hat{Y}_t = 1 + \underline{f}_t\ \underline{K}_t\ \underline{f}_t'$$

$$\underline{A}_t = \underline{K}_t\ \underline{f}_t'\ (\ \hat{Y}_t\ )^{-1}$$

$$\underline{Q}_t^{-1} = (\ \underline{I} - \underline{A}_t\ \underline{f}_t\ )\ \underline{K}_t$$

$$\underline{m}_t = \underline{m}_{t-1} + \underline{A}_t\ e_t$$

where $e_t = y_t - \underline{f}_t\ \underline{m}_{t-1}\ .$

From these recurrence relations it will be noticed that no matrix inversion is required. once we set the prior matrix $\underline{Q}_0^{-1}$ in some convenient way, say

$$\underline{Q}_0^{-1} = c\,\underline{I} \quad ;$$

where c is any constant greater than zero, $\underline{Q}_t^{-1}$ is updated automatically through the given recurrence relations. For the whole operation we just need to provide the prior information on $\underline{m}_0$ and $\underline{Q}_o^{-1}$ to the dynamic system, where $\underline{m}_0$ is an estimate of the parameter $\underline{\theta}$ at time $t = 0$.

The discount factor β associated with the covariance matrix $\underline{K}_t$ helps to avoid the covariance wind-up (in analogy with integrator wind-up in simple regulators),i.e. due to β the matrix $\underline{K}_t$ grows exponentially and does not approach to zero matrix. For more discussion see Astrom (1981).

3.3.2 Time Series Form of D.E.W.R.

In a time series case we adopt the usual moving origin representation so that at time t

$$Y_{t+i} = \underline{f}_i\,\underline{\theta} + \delta_{t+i} \quad ; \quad \delta_{t+i} \sim N(0,\ \sigma_\delta^2) \qquad (3.3.2.1)$$

$$\underline{f}_i = \underline{f}_0\,\underline{G}^i \ ,$$

$$(\ \underline{f}_0^{'}\ ,\ \underline{f}_1^{'}\ ,\ \ldots\ldots,\ \underline{f}_{n-1}^{'}) \quad \text{is of rank } n$$

and the transition matrix $\underline{G}$ has non zero eigenvalues which are not necessarily distinct.

Following the definitions (2.3.1.1) - (2.3.1.4) given in corollary 1 of the theorem (T1) of E.W.R., the recurrence relations developed for the ordinary case of D.E.W.R. can easily be extended to the time series case.

Considering $\underline{f} = \underline{f}_O$ the required recurrence relations for updating $\underline{m}_t$ are

$$\underline{K}_t = \underline{G}\ \underline{Q}_{t-1}^{-1}\ \underline{G}^{'}\ /\ \beta \qquad (3.3.2.2)$$

$$\hat{Y}_t = 1 + \underline{f} \; \underline{K}_t \; \underline{f}' \tag{3.3.2.3}$$

$$\underline{A}_t = \underline{K}_t \; \underline{f}' \; (\hat{Y}_t)^{-1} \tag{3.3.2.4}$$

$$\underline{Q}_t^{-1} = (\underline{I} - \underline{A}_t \; \underline{f} \;) \; \underline{K}_t \tag{3.3.2.5}$$

$$e_t = y_t - \underline{f} \; \underline{G} \; \underline{m}_{t-1} \tag{3.3.2.6}$$

$$\underline{m}_t = \underline{G} \; \underline{m}_{t-1} + \underline{A}_t \; e_t \tag{3.3.2.7}$$

The k-steps ahead forecast function in this case is

$$\begin{aligned} F_t(k) &= E (Y_{t+k} \mid D_t) \\ &= \underline{f} \; \underline{G}^k \; \underline{m}_t \end{aligned} \tag{3.3.2.8}$$

where D_t has its usual meaning.

Restricting β by $0 < \beta < \min_i |\lambda_i^2|$

the limits as $t \to \infty$

$$\underline{Q}_t \to \underline{Q} = \sum_{i=0}^{\infty} \beta^i \; (\underline{G}')^{-i} \; \underline{f}' \; \underline{f} \; \underline{G}^{-i} \, , \tag{3.3.2.9}$$

$$\underline{A}_t \to \underline{A} = \underline{Q}^{-1} \; \underline{f}$$

and

$$\lim_{t \to \infty} \{ \prod_{i=1}^{n} (1 - \lambda_i B) \; y_t - \prod_{i=1}^{n} (1 - \beta B / \lambda_i) \; e_t \} = 0 \tag{3.3.2.10}$$

exist.

The result (3.3.2.10) is a special case of the G.E.W.R. theorem(T3) given in section (5.3.3). The stated restriction on the discount factor β is severe so that E.W.R. and D.E.W.R. are only worth considering if $\min_r |\lambda_i|$ is close to 1. If $e_t \sim (o, \sigma^2)$ are independent, $|\lambda_i| < 1$ for $i = 1,\ldots,p$, and $\lambda_i = 1$ for $i = p+1,\ldots,n$, then the series Y_t can be represented as an ARIMA(p,n-p,n) process with parametric parsimony (2) if λ_i are all specified and parametric parsimony (n+2) if none of the λ_i are known. For related discussion see Harrison-Akram (1983).

The dynamic system developed for E.W.R. models is based on the assumption that the noise present in the time series is white . In the forthcoming chapters five and six, this dynamic system is developed further for time series driven by ARMA type Coloured Noise processes.

3.4 Dynamic Linear Models (D.L.M.)

In 1976 Harrison-Stevens introduced a wide class of Dynamic Linear Models System based on Bayesian principles. A univariate form of this dynamic system is considered here.

The general dynamics of the discrete time variant D.L.M., characterised by the quadruple

$$(\underline{f} , \underline{G} , V , \underline{W})_t$$

is expressed in the form of Markov or first order difference equations as

$$Y_t = \underline{f} \; \underline{\theta}_t + v_t \qquad (3.4.0.1)$$

$$\underline{\theta}_t = \underline{G} \; \underline{\theta}_{t-1} + \underline{w}_t \qquad (3.4.0.2)$$

where (3.4.0.1) is an observation equation and (3.4.0.2) is a system equation, such that at time t Y_t is an observation, $\underline{f}$ is a (1xn) vector of some known functions, $\underline{\theta}_t$ is an (nx1) state vector or vector of stochastic parameters to be estimated and $\underline{G}$ is a transition matrix of full rank. v_t is a Guassian White Noise and $\underline{w}_t$ is a (nx1) Gaussian White Noise vector, such that

$$E (v_t) = 0 \quad , \quad E (v_t^2) = V$$

$$E (\underline{w}_t) = \underline{0} \quad , \quad E (\underline{w}_t \; \underline{w}_t') = \underline{W}$$

and v_t is independent of $\underline{w}_t$.

Defining D_{t-1} as the data history of time series y_{t-1} , y_{t-2} , ..., y_1 known at time t-1, i.e. prior to the arrival of the observation at time t, $D_t=(y_t , D_{t-1})$ and expressing probabilistic information in the Bayesian form as:

i) $(\underline{\theta}_{t-1} | D_{t-1}) \sim N(\underline{m}_{t-1} ; \underline{C}_{t-1})$,

where

$$E(\underline{\theta}_{t-1} | D_{t-1}) = \underline{m}_{t-1}$$

and $\quad Var(\underline{\theta}_{t-1} | D_{t-1}) = E(\underline{\theta}_{t-1} - \underline{m}_{t-1})(\underline{\theta}_{t-1} - \underline{m}_{t-1})'$

$$= \underline{C}_{t-1}.$$

ii) Relating the information D_{t-1} at time t-1 to the state of the system at time t as

$$(\underline{\theta}_t | D_{t-1}) \sim N(\hat{\underline{\theta}}_t ; \underline{R}_t)$$

where $\hat{\underline{\theta}}_t$ is a predictor state estimator, such that

$$\hat{\underline{\theta}}_t = E(\underline{\theta}_t | D_{t-1})$$

$$= \underline{G}\, \underline{m}_{t-1} \quad \text{and}$$

$\underline{R}_t$ is the predicted estimation error covariance matrix, such that

$$\underline{R}_t = E\{(\underline{\theta}_t - \hat{\underline{\theta}}_t)(\underline{\theta}_t - \hat{\underline{\theta}}_t)' | D_{t-1}\}$$

$$= \underline{G}\, \underline{C}_{t-1} \underline{G}' + \underline{W}$$

iii) $(y_t | D_{t-1}) \sim N(\hat{y}_t ; \hat{Y}_t)$

where $\hat{y}_t$ is a one step ahead forecast, such that

$$\hat{y}_t = E(y_t | D_{t-1}) = \underline{f}\ \underline{G}\, \underline{m}_{t-1}$$

$\hat{Y}_t$ is the variance of the forecast errors, such that

$$\hat{Y}_t = E\{(y_t - \hat{y}_t)^2 | D_{t-1}\}$$

under the assumption that the one step ahead forecast errors form a Gaussian White Noise sequence.

iv) $(\underline{\theta}_t | D_t) \sim N(\underline{m}_t ; \underline{C}_t)$

where $\underline{m}_t$ is the corrected or updated state estimator and $\underline{C}_t$ is the corrected or updated covariance matrix of the state estimator, then the posterior $\underline{m}_t$ and $\underline{C}_t$ are calculated through the recurrence relations

$$\underline{R}_t = \underline{G}\ \underline{C}_{t-1}\ \underline{G}' + \underline{W} \tag{3.4.0.3}$$

$$\hat{Y}_t = V + \underline{f}\ \underline{R}_t\ \underline{f}' \tag{3.4.0.4}$$

$$\underline{A}_t = \underline{R}_t\ \underline{f}'\ (\hat{Y}_t)^{-1} \tag{3.4.0.5}$$

$$\underline{C}_t = (\underline{I} - \underline{A}_t\ \underline{f})\ \underline{R}_t \tag{3.4.0.6}$$

$$F_t(1) = \underline{f}\ \underline{G}\ \underline{m}_t \tag{3.4.0.7}$$

$$e_t = y_t - \underline{f}\ \underline{G}\ \underline{m}_{t-1} \tag{3.4.0.8}$$

$$\underline{m}_t = \underline{G}\ \underline{m}_{t-1} + \underline{A}_t\ e_t \tag{3.4.0.9}$$

and the k-steps ahead forecast function at time t is given as

$$F_t(k) = \underline{f}\ \underline{G}^k\ \underline{m}_t \tag{3.4.0.10}$$

The prior information required for these recurrence relations is the setting of $\underline{m}_0$ and $\underline{C}_0$. The $\underline{W}$ matrix may be chosen or specified in various ways (see Harrison-Stevens(1976)), one of which is given in section (3.4.2). If the variance V is unknown and or it is time variant, it is estimated using the on-line Bayesian learning procedure explained in section (3.4.1)

The Dynamic Linear Model representation above provides quite a general framework for the analysis and forecasting of univariate time series for many stochastic phenomena.

Generally the .D.L.M. system is expressed in a canonical form. If any other form or reparameterisation is required, this can be achieved following the transformations and Similar Models technique given in section (6.6).

3.4.1 On-Line Variance Learning

In Dynamic Linear Models (D.L.M.) if the variance V is not known at time t then it is estimated on-line through a Bayesian learning process in a recursive manner. The recurrence relations in this respect are:

$$V_t = X_t / N_t \qquad (3.4.1.1)$$

$$X_t = \beta_v X_{t-1} + (1 - \underline{f} \ \underline{A}_t) e_t^2 \qquad (3.4.1.2)$$

$$N_t = \beta_v N_{t-1} + 1 \qquad (3.4.1.3)$$

where β_v is a discount factor associated with the variance V_t and $0 < \beta_v < 1$.

This is a special case of a variance learning procedure (6.8) for Coloured Noise processes, described in chapter six.

For robustness or protection from the outliers, the equation (3.4.1.2) may be modified as

$$X_t = \beta_v X_{t-1} + (1 - \underline{f} \ \underline{A}_t) d_t \qquad (3.4.1.4)$$

where $$d_t = \min (e_t^2 ; \xi \hat{Y}_t) \qquad (3.4.1.5)$$

and ξ is a confidence factor. Usually $\xi = 4$ and $\beta_v \geq 0.96$ is considered.

For the first n+1 (where n is the order of the model) observations, instead of using the estimated value of variance V_t an initial estimate V_0 is used as for these observations no significant contribution to the variance learning is expected. For more discussion see section (6.8) of chapter six.

The dynamic Bayesian learning procedure introduced, improves the performance of the Dynamic Linear Models quite significantly as very rarely in real life cases is the variance V_t constant.

3.4.2 W - Matrix

There are various techniques for setting a constant $\underline{W}$ matrix of Dynamic Linear Models (D.L.M.) which has limiting E.W.R. Forecasting functions. One of these is described here with special reference to a canonical form of a dynamic system, where, $\underline{W} = \text{diag}(W_1, \ldots, W_n)$. For any other form of a dynamic system the relevant $\underline{W}$ matrix may be found by following the method of transformation and similar models explained in section (6.6) of chapter six.

3.4.2.1 First Order (n=1) Case

For a first order model

$$Y_t = \theta_t + v_t \quad ; v_t \sim N(0, V) \qquad (3.4.2.1.1)$$

$$\theta_t = \lambda\theta_{t-1} + w_t \quad ; w_t \sim N(0, W) \qquad (3.4.2.1.2)$$

the W, a scalar, can be obtained following the procedure described below.

Considering backward shift operator B such that $B\theta_t = \theta_{t-1}$ we can write the parameter equation (3.4.2.1.2) as

$$(1-\lambda B)\theta_t = w_t \ . \qquad (3.4.2.1.3)$$

Writing an observation y_t from the observation equation Y_t as

$$y_t = \theta_t + v_t$$

and multiplying both sides by $(1-\lambda B)$ we get

$$(1-\lambda B)y_t = (1-\lambda B)\theta_t + (1-\lambda B)v_t .$$

Substituting (3.4.2.1.3) for $(1- \lambda B)\ \theta_t$ we get

$$(1-\lambda B)y_t = w_t + (1-\lambda B)v_t \ . \qquad (3.4.2.1.4)$$

Now from (3.3.2.10) we know that for n=1 and $\lambda_i = \lambda$

$$x_t = (1-\lambda B)y_t \tag{3.4.2.1.5}$$

$$= (1-\beta B/\lambda)e_t. \tag{3.4.2.1.6}$$

Substituting (3.4.2.1.4) in (3.4.2.1.5) we get

$$x_t = w_t + (1-\lambda B)v_t = (v_t + w_t) - \lambda B v_t \tag{3.4.2.1.7}$$

Writing $c_1 = v_t + w_t$ and $c_2 = -\lambda v_t$ and defining the Autocovariance Generating Function (A.C.G.F.) of x_t by

$$\gamma_x = E\,\{(c_1 + Bc_2)(c_1 + B^{-1}c_2)\}$$

we get the autocovariances $\gamma_k = \gamma_{-k}$ at lags k=0,1

$$\gamma_1 = E(c_1 c_2) = -\lambda V \tag{3.4.2.1.8}$$

$$\gamma_0 = E(c_1^2 + c_2^2) = (1+\lambda^2)V + W\,. \tag{3.4.2.1.9}$$

Also from (3.4.2.6)

$$\gamma_x = E\{\,(1-\beta B/\lambda)(1-\beta B^{-1}/\lambda)\,\}e_t^2$$

we get the autocovariances

$$\gamma_1 = -(\beta/\lambda)\,\hat{Y} \tag{3.4.2.1.10}$$

$$\gamma_0 = (1+\beta^2/\lambda^2)\,\hat{Y} \tag{3.4.2.1.11}$$

where in the limit $t\to\infty$ $E(e_t^2) = \hat{Y}$.
Comparing the respective autocovariances we get

$$\hat{Y} = (\lambda^2/\beta)V \tag{3.4.2.1.12}$$

and $$W = (\lambda^2-\beta)(1-\beta)V/\beta \tag{3.4.2.1.13}$$

the required W.

As a special case when $\lambda = 1$ the result reduces to

$$W = (1-\beta)^2 V/\beta\,. \tag{3.4.2.1.14}$$

3.4.2.2 Second Order (n=2) Case

For a second order model

$$Y_t = (1\ 0)\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}_t + v_t \quad ; \quad v_t \sim N(0, V) \qquad (3.4.2.2.1)$$

$$= \begin{bmatrix}\lambda_1 & 1\\ 0 & \lambda_2\end{bmatrix}\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}_{t-1} + \begin{pmatrix}w_1\\ w_2\end{pmatrix}_t ; \begin{pmatrix}w_1\\ w_2\end{pmatrix}_t \sim N(\underline{0}, \underline{W})$$

(3.4.2.2.2)

the matrix $\underline{W} = \mathrm{diag}(\ W_1,\ W_2\ )$ is obtained following the procedure of first order case.

Using the backward shift operator B, the parameter equation (3.4.2.2.2) may be written as

$$(1-\lambda_1 B)\theta_{1,t} = \theta_{2,t-1} + w_{1,t} \qquad (3.4.2.2.3)$$

$$(1-\lambda_2 B)\theta_{2,t} = w_{2,t} \qquad (3.4.2.2.4)$$

Multiplying equation (3.4.2.2.3) by $(1-\lambda_2 B)$ and equation (3.4.2.2.4) by B we get

$$\prod_{i=1}^{2}(1-\lambda_i B)\theta_{1,t} = (1-\lambda_2 B)(\theta_{2,t-1} + w_{1,t})$$

$$(1-\lambda_2 B)\theta_{2,t} = Bw_{2,t}\ .$$

Substituting the second equation in to the first equation we get

$$\prod_{i=1}^{2}(1-\lambda_i B)\theta_{1,t} = Bw_{2,t} + (1-\lambda_2 B)w_{1,t}. \qquad (3.4.2.2.5)$$

Writing an observation y_t from the above observation equation of Y_t as

$$y_t = \theta_{1,t} + v_t$$

and multiplying both sides by $\prod_{i=1}^{2}(1-\lambda_i B)$ we get

$$\prod_{i=1}^{2}(1-\lambda_i B)y_t = \prod_{i=1}^{2}(1-\lambda_i B)(\theta_{1,t} + v_t\ ).$$

Substituting here (3.4.2.2.5) we get

$$\prod_{i=1}^{2}(1-\lambda_i B)y_t = Bw_{2,t} + (1-\lambda_2 B)w_{1,t} + \prod_{i=1}^{2}(1-\lambda_i B)v_t .$$

Re-arranging the terms in B, we get

$$\prod_{i=1}^{2}(1-\lambda_i B)y_t = (w_1+v)_t + \{w_2-\lambda_2 w_1-(\lambda_1+\lambda_2)v\}_t B + \lambda_1\lambda_2 B^2 v_t . \qquad (3.4.2.2.6)$$

Now from the corollary 2 of G.E.W.R. theorem (5.3.3) the limiting result for n=2 is

$$x_t = \prod_{i=1}^{2}(1-\lambda_i B)y_t \qquad (3.4.2.2.7)$$

$$= \prod_{i=1}^{2}(1-\beta B/\lambda_i)e_t . \qquad (3.4.2.2.8)$$

Substituting (3.4.2.2.6) in (3.4.2.2.7) we get

$$x_t = (w_1+v)_t + B\{w_2 - \lambda_2 w_1 - (\lambda_1+\lambda_2)v\}_t + \lambda_1\lambda_2 B^2 v_t$$

$$(3.4.2.2.9)$$

Letting $c_1 = (w_1+v)_t$, $c_2 = \{(w_2-\lambda_2 w_1-(\lambda_1+\lambda_2)v\}_t$ and $c_3 = \lambda_1\lambda_2 v_t$ and defining A.C.G.F.

$$\gamma_x = E\{ (c_1+Bc_2+B^2c_3)(c_1+B^{-1}c_2+B^{-2}c_3) \}$$

we get the autocovariances $\gamma_k = \gamma_{-k}$ at lag k = 0,1,2 as

$$\gamma_0 = E(c_1^2+c_2^2+c_3^2)$$

$$= \{1+(\lambda_1+\lambda_2)^2 + \lambda_1^2\lambda_2^2\}V + (1+\lambda_2^2)W_1 + W_2 \qquad (3.4.2.2.10)$$

$$\gamma_1 = E\{c_2(c_1+c_3)\} = -\lambda_2 W_1 - (1+\lambda_1\lambda_2)(\lambda_1+\lambda_2)V \qquad (3.4.2.2.11)$$

$$\gamma_2 = E(c_1c_3) = \lambda_1\lambda_2 V . \qquad (3.4.2.2.12)$$

Similarly writing the A.C.G.F. of (3.4.2.2.8), i.e.

$$\gamma_x = E\ \{ \prod_{i=1}^{2} (1-\beta B/\lambda_i)(1-\beta B^{-1}/\lambda_i) \}\ e_t^2 \tag{3.4.2.2.13}$$

and writing in limit $t\to\infty$ $E(e_t^2) = \hat{Y}$, we get the autocovariances at lag $k = 0,1,2$ as

$$\gamma_0 = \beta^2\ \hat{Y}\ \{ (\lambda_1+\lambda_2)^2 + (\lambda_1\lambda_2/\beta)^2 + \beta^2\}/\lambda_1^2\lambda_2^2 \tag{3.4.2.2.14}$$

$$\gamma_1 = -\ \beta\hat{Y}\ \{(\lambda_1\lambda_2 + \beta^2)(\lambda_1+\lambda_2)\ \}\ /\lambda_1^2\lambda_2^2 \tag{3.4.2.2.15}$$

$$\gamma_2 = \beta^2\hat{Y}/\lambda_1\lambda_2 \tag{3.4.2.2.16}$$

Comparing the respective autocovariances we get

$$\hat{Y} = \prod_{i=1}^{2} (\lambda_i^2\ /\beta) \tag{3.4.2.2.17}$$

$$W_1 = (1-\beta)(\lambda_1+\lambda_2)(\lambda_1\lambda_2 - \beta)V/(\lambda_2\beta) \tag{3.4.2.2.18}$$

$$W_2 = -(1+\lambda_2^2)W_1 + (1-\beta^2)(\lambda_1^2\lambda_2^2 - \beta^2)V/\beta^2$$

$$= (1-\beta)(\lambda_1\lambda_2-\beta)(\lambda_1-\lambda_2\beta)(\lambda_2^2 - \beta)V/(\lambda_2\beta^2) \tag{3.4.2.2.19}$$

the required result.

<u>Special Cases</u>

i) If $\lambda_1 = 1$ and $\lambda_2 = \lambda$ such that $\underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & \lambda \end{bmatrix}$ then the above results reduce to

$$W_1 = (1-\beta)(1+\lambda)(\lambda-\beta)V/(\lambda\beta) \tag{3.4.2.2.20}$$

$$W_2 = (1-\beta)(\lambda-\beta)(1-\lambda\beta)(\lambda^2-\beta)V/(\lambda\beta^2) \tag{3.4.2.2.21}$$

ii) If $\lambda_1 = \lambda_2 = \lambda$ such that $\underline{G} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$ then we can write

$$W_k = \binom{2}{k}\ \{ (\lambda^2 - \beta)(1-\beta)/\beta\}^k V \tag{3.4.2.2.22}$$

for $k = 1,2$.

iii) If $\lambda_1 = \lambda_2 = 1$ such that $\underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ then the results (3.4.2.2.18) and (3.4.2.2.19) can be written as

$$W_k = \binom{2}{k} \{ (1-\beta)^2/\beta \}^k V \qquad (3.4.2.2.23)$$

for k = 1,2.

<u>Comments</u> :

i) The setting of the $\underline{W}$ matrix is not unique. However, many possible settings can be interpreted simply as time shifts of the parameters. For example, the second order polynomial E.W.R. model

$$y_t = \mu_t + v_t \qquad ; \quad v_t \sim N(0, V)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + w_{1,t} \qquad ; \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}_t \sim N(\underline{0}, \underline{W})$$

$$\beta_t = \beta_{t-1} + w_{2,t}$$

may have

$$\underline{W} = \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \qquad \text{or} \qquad \underline{W} = \begin{bmatrix} W_1 + W_2 & W_2 \\ W_2 & W_2 \end{bmatrix}$$

where each setting gives identical forecasts. The growth parameter β_t in the first case is β_{t+1} for the second case (i.e. a time shift of 1 interval). Here W_1 and W_2 are as defined above in case (iii). The many time shifts, each correspond to a

$$\underline{W} \in \{ \underline{W} = \begin{bmatrix} W_1 + aW_2 & a\underline{W}_2 \\ a\underline{W}_2 & \underline{W}_2 \end{bmatrix} \quad ; \ (2a - 1)^2 < 1 + 4W_1/W_2 \}$$

giving identical forecast distributions and having the same limiting E.W.R. forecast function (see Harrison-Akram (1983)).

ii) The procedure for deriving the $\underline{W}$ matrix has been given up to a second order case. For higher order cases, the $\underline{W}$ matrices can be easily found following the procedure described above.

3.4.3 Special Cases of Dynamic Linear Models (D.L.M.)

3.4.3.1 First Order D.L.M.

This is a special case of the D.L.M. (3.4) when n=1. In this case we consider

$$\underline{f} = 1 \quad , \quad \underline{G} = \lambda \quad , \quad \underline{w}_t = w_t \quad \text{and} \quad \underline{\theta}_t = \theta_t .$$

The observation and parameter equations in this case are

$$Y_t = \theta_t + v_t \quad ; \quad v_t \sim N(0, V)$$

$$\theta_t = \lambda\theta_{t-1} + w_t \quad ; \quad w_t \sim N(0, W). \qquad (3.4.3.1.1)$$

Given the prior

$$(\theta_{t-1} | D_{t-1}) \sim N(m_{t-1} , C_{t-1})$$

the posterior estimate m_t of θ_t such that

$$(\theta_t | D_t) \sim N(m_t , C_t)$$

is recursively obtained as

$$R_t = \lambda^2 C_{t-1} + W$$

$$\hat{Y}_t = \lambda^2 C_{t-1} + W + V$$

$$A_t = (\lambda^2 C_{t-1} + W)/(\lambda^2 C_{t-1} + W + V)$$

$$C_t = A_t \; V_t$$

$$e_t = y_t - \lambda \; m_{t-1}$$

$$m_t = \lambda \; m_{t-1} + A_t e_t \qquad (3.4.3.1.2)$$

The k-steps ahead forecast function for this model is

$$F_t(k) = \lambda^k m_t \qquad (3.4.3.1.3)$$

Limiting Results

In the limit $t \to \infty$, $V_t \to V$,

$A_t \to A = (\lambda^2 C + W)/(\lambda^2 C + W + V)$

$C_t \to C = A\,V.$

Substituting the value of C in A and simplifying we get

$$W = (1 + \lambda^2 A - \lambda^2)AV/(1-A). \qquad (3.4.3.1.4)$$

In the previous section (3.4.2.1) we derived the value of W as

$$W = (\lambda^2 - \beta)(1 - \beta)V/\beta \qquad (3.4.3.1.5)$$

Comparing the W values we get on simplification

$$A = (\lambda^2 - \beta)/\lambda^2 \qquad (3.4.3.1.6)$$

Substituting this value of A in the above limiting expression of C we get

$C = (\lambda^2 - \beta)V/\lambda^2$

or $V = \lambda^2 C/(\lambda^2 - \beta)$.

Substituting this value of V in (3.4.3.1.5) we get a relationship between W and C, i.e.

$$W = (1 - \beta)\lambda^2 C/\beta \qquad (3.4.3.1.7)$$

From the limiting result for $\underline{Q}$ (2.3.3.1) we know that

$Q = \lambda^2/(\lambda^2 - \beta)$ or $Q^{-1} = (\lambda^2 - \beta)/\lambda^2$

for $\underline{f} = 1$ and $\underline{G} = \lambda$

Comparing Q^{-1} with C we see that $C = Q^{-1} V$.
Substituting this value of C in (3.4.3.1.7) we get

$$W = (1-\beta) Q^{-1} V\beta \quad (3.4.3.1.8)$$

This relation helps us to understand the link between D.L.M. and E.W.R. models, such as D.E.W.R. (3.3.2) models. In the limit both, the first order D.L.M. (3.4.3.1.1) and the first order D.E.W.R. (3.3.2) yield the same forecasts. In both cases the limiting value of the updating coefficient A is equal to (3.4.3.1.6) and in the limit $t \to \infty$

$$m_t = \lambda m_{t-1} + A e_t$$

$$= \lambda m_{t-1} + \{(\lambda^2 - \beta)/\lambda^2\} e_t \quad (3.4.3.1.9)$$

Further if $\lambda = 1$, the first order D.L.M (3.4.3.1.1) reduces to a model called the Steady Model. In such a case the values of W and A reduce to

$$W = (1-\beta)^2 V/\beta$$

and $\quad A = 1-\beta \quad .$

The above result for m_t consequently reduces to

$$m_t = m_{t-1} + (1-\beta) e_t \quad (3.4.3.1.10)$$

This result is the same as (2.2.0.7), i.e. the simple exponential smoothing result given in chapter two.

The above results reveal that the particular setting of W (3.4.3.1.5) establishes a close link between the first order D.L.M. and the first order E.W.R. models.

For second order models such a link is described in the forthcoming sub-section.

3.4.3.2 Second Order Dynamic Linear Model

This is a special case of the general form of D.L.M. defined by observation equation (3.4.0.1) and system equation (3.4.0.2) when n = 2,

$$\underline{f} = (1 \quad 0), \quad \underline{G} = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{bmatrix}, \quad \underline{\theta}_t = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}_t \quad \text{and} \quad \underline{w}_t = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}_t$$

The model in this case becomes

$$Y_t = \theta_{1,t} + v_t$$

$$\theta_{1,t} = \lambda_1 \theta_{1,t-1} + \theta_{2,t-1} + w_{1,t}$$

$$\theta_{2,t} = \lambda_2 \theta_{2,t-1} + w_{2,t} \qquad (3.4.3.2.1)$$

For data $D_{t-1} = y_{t-1}, \ldots, y_1$ and $D_t = (y_t, D_{t-1})$, given the prior

$$(\underline{\theta}_{t-1} | D_{t-1}) \sim N(\underline{m}_{t-1}, \underline{C}_{t-1})$$

and setting the $\underline{W}$ matrix either as (3.4.2.2) or in any other appropriate form, the posterior estimate $\underline{m}_t$ of $\underline{\theta}_t$ such that

$$(\underline{\theta}_t | D_t) \sim N(\underline{m}_t, \underline{C}_t)$$

is obtained recursively using the recurrence relations (3.4.0.3) to (3.4.0.10).

The variance V if is not constant and is unknown at time t, is estimated on line through the Bayesian learning process presented in section (3.4.1).

Forecast Function

The k-steps ahead forecast function for the second order model (3.4.3.2.1) is

$$F_t(k) = E(Y_{t+k} | D_t) = \lambda_1^k m_{1,t} + \lambda_1^{k-1} \cdot \sum_{i=0}^{k-1} (\lambda_2/\lambda_1)^i m_{2,t}$$

for all finite $k \geq 1$. $\qquad (3.4.3.2.2)$

Limiting Results

When $t \to \infty$, the limits

$$V_t \to V, \quad \hat{Y}_t \to \hat{Y}, \quad \underline{A}_t \to \underline{A}, \quad \underline{C}_t \to \underline{C} \text{ and}$$

$$\underline{m}_t = \underline{G}\,\underline{m}_{t-1} + \underline{A}\,e_t$$

exist . However, in general, the precise expression for $\underline{A}$ is not known analytically, except for a particular setting of the $\underline{W}$ matrix. For example in the present second order case if we set

$$\underline{W} = \begin{bmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{bmatrix} \tag{3.4.3.2.3}$$

by defining

$$\underline{W} = \{(1-\beta)/\beta\}\,\underline{G}\,\underline{C}\,\underline{G}'$$

$$= \{(1-\beta)/\beta\}\,\underline{G}\,\underline{Q}^{-1}V\,\underline{G}' \tag{3.4.3.2.4}$$

where $\underline{Q}$ is the limiting value of $\underline{Q}_t$ defined for E.W.R. models by (2.3.3.1), i.e.

$$\underline{Q} = \underline{f}'\underline{f} + \beta\,(\underline{G}')^{-1}\,\underline{Q}\,\underline{G}^{-1}$$

the inverse of which for a second order E.W.R. model is (2.3.3.5), then the limiting value of $\underline{A}_t = \underline{A} = (A_1 \quad A_2)'$ is as derived in section (2.3.3) of chapter two. The elements of $\underline{W}$ obtained in the present case, using the above expression (3.4.3.2.4) are

$$W_{11} = (1-\beta)(\lambda_1^2\lambda_2^2 - \beta^2)/\beta^2$$

$$W_{12} = \lambda_1(1-\beta)(\lambda_1\lambda_2 - \beta)(\lambda_2^2 - \beta)/\beta^2$$

$$W_{22} = (1-\beta)(\lambda_1\lambda_2 - \beta)^2(\lambda_2^2 - \beta)/\beta^2; \tag{3.4.3.2.5}$$

and the limiting value of the updating vector $\underline{A}_t = \underline{A} = (A_1 \quad A_2)'$

$$A_1 = (\lambda_1^2\lambda_2^2 - \beta^2)/\lambda_1^2\lambda_2^2 \text{ and } A_2 = (\lambda_1\lambda_2 - \beta)(\lambda_2^2 - \beta)/\lambda_1\lambda_2^2 \tag{3.4.3.2.6}$$

Other special cases of interest are:

Limiting Results

When $t \to \infty$, the limits

$$V_t \to V, \quad \hat{Y}_t \to \hat{Y}, \quad \underline{A}_t \to \underline{A}, \quad \underline{C}_t \to \underline{C} \text{ and}$$

$$\underline{m}_t = \underline{G}\,\underline{m}_{t-1} + \underline{A}\,e_t$$

exist . However, in general, the precise expression for $\underline{A}$ is not known analytically, except for a particular setting of the $\underline{W}$ matrix. For example in the present second order case if we set

$$\underline{W} = \begin{bmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{bmatrix} \qquad (3.4.3.2.3)$$

by defining

$$\underline{W} = \{(1-\beta)/\beta\}\,\underline{G}\,\underline{C}\,\underline{G}'$$

$$= \{(1-\beta)/\beta\}\,\underline{G}\,\underline{Q}^{-1}V\,\underline{G}' \qquad (3.4.3.2.4)$$

where $\underline{Q}$ is the limiting value of $\underline{Q}_t$ defined for E.W.R. models by (2.3.3.1), i.e.

$$\underline{Q} = \underline{f}'\underline{f} + \beta\,(\underline{G}')^{-1}\,\underline{Q}\,\underline{G}^{-1}$$

the inverse of which for a second order E.W.R. model is (2.3.3.5), then the limiting value of $\underline{A}_t = \underline{A} = (A_1 \quad A_2)'$ is as derived in section (2.3.3) of chapter two. The elements of $\underline{W}$ obtained in the present case, using the above expression (3.4.3.2.4) are

$$W_{11} = (1-\beta)(\lambda_1^2\lambda_2^2 - \beta^2)/\beta^2$$

$$W_{12} = \lambda_1(1-\beta)(\lambda_1\lambda_2 - \beta)(\lambda_2^2 - \beta)/\beta^2$$

$$W_{22} = (1-\beta)(\lambda_1\lambda_2 - \beta)^2(\lambda_2^2 - \beta)/\beta^2; \qquad (3.4.3.2.5)$$

and the limiting value of the updating vector $\underline{A}_t = \underline{A} = (A_1 \quad A_2)'$

$$A_1 = (\lambda_1^2\lambda_2^2 - \beta^2)/\lambda_1^2\lambda_2^2 \text{ and } A_2 = (\lambda_1\lambda_2 - \beta)(\lambda_2^2 - \beta)/\lambda_1\lambda_2^2$$

$$(3.4.3.2.6)$$

Other special cases of interest are:

i) If $\lambda_1 = \lambda_2 = \lambda$,then the limiting value of $\underline{A}_t$ is given by (2.3.3.9) with

$$A_1 = (\lambda^4 - \beta^2)/\lambda^4 \quad \text{and} \quad A_2 = (\lambda^2 - \beta)^2/\lambda^3 \qquad (3.4.3.2.7)$$

ii) If $\lambda_1 = \lambda_2 = 1$ then $\underline{A}$ is given by (2.3.3.12) with

$$A_1 = 1-\beta^2 \qquad \text{and} \quad A_2 = (1-\beta)^2 \qquad (3.4.3.2.8)$$

iii) If $\lambda_1 = 1, \lambda_2 = \lambda$, then $\underline{A}$ is given by (2.3.3.15) with

$$A_1 = (\lambda^2 - \beta^2)/\lambda^2 \quad \text{and} \quad A_2 = (\lambda-\beta)(\lambda^2-\beta)/\lambda^2 \qquad (3.4.3.2.9)$$

Comments

i) The above results show that if the $\underline{W}$ matrix is set according to (3.4.3.2.4), then in the limit $t \to \infty$ the second order D.L.M. yields forecasts equivalent to the second order E.W.R. and D.E.W.R. models given in section (2.3) of chapter two and section (3.3.2) of chapter three respectively, when similar settings of $\underline{f}$ and $\underline{G}$ are considered. This is also true for higher order cases.

ii) The second order D.L.M. is useful in practice, especially the above case (iii) due to its asymptotic behaviour. Second order models are commonly used to represent a trend or a low frequency. These models often give quite satisfactory results, especially when dealing with economic and industrial time series.

Moreover analogous to Taylor Series, it is well known that in most of the cases a time series trajectory can be reasonably represented by the first two terms of the Taylor Series

$$f(y) = \sum_{i=0}^{\infty} f^{(i)}(m)(y-m)^i/i!.$$

If m is a reasonable estimate of y, then

$$f(y) \cong f(m) + f^{(1)}(m)(y-m)$$

as the contribution of higher order terms in (y-m) become insignificant.

3.5 E.W.R. and Dynamic Linear Models

3.5.1 Polynomial E.W.R.

From corollary 4 (5.3.4.3) of the G.E.W.R. theorem (T3) for $\lambda = 1$

$$(1 - B)^n y_t = (1 - \beta B)^n e_t \tag{3.5.1.1}$$

If the e_t are independent $(0, \sigma^2)$ then the series Y_t can be represented as an ARIMA(0,n,n) process with parsimony (2).

For D.L.M. modellers using Bayesian Forecasting, the following constant D.L.M of parsimony (2) has a limiting polynomial forecast function equivalent to that of E.W.R. with discount factor β (see Harrison-Akram(1983)).

$$Y_t = \underline{f}\,\underline{\theta}_t + v_t \qquad ; v_t \sim N(0, V)$$

$$\underline{\theta}_t = \underline{J}_n(1)\,\underline{\theta}_{t-1} + w_t \qquad ; w_t \sim N(0, \underline{W}) \tag{3.5.1.2}$$

where

$\underline{f} = (1,0,\ldots,0)$, $\underline{J}_n(1)$ is a (nxn) matrix of Jordan form with n unit eigenvalues,

$\underline{W} = \operatorname{diag}(W_1, \ldots, W_n)$,

$$W_k = \binom{n}{k} \{(1-\beta)^2/\beta\}^k \qquad \text{for } k = 1,\ldots,n. \tag{3.5.1.3}$$

$$\lim_{t\to\infty} V(e_t) = \lim_{t\to\infty} \hat{Y}_t = \hat{Y} = V/\beta^n \tag{3.5.1.4}$$

and

$$E(\underline{\theta}_t | D_t) = \underline{m}_t = \underline{J}_n(1)\,\underline{m}_{t-1} + \underline{A}_t e_t \tag{3.5.1.5}$$

with $\lim_{t\to\infty} \underline{A}_t = \underline{A} = (A_1, \ldots, A_n)'$

where, defining $A_1 = 1 - \beta^n$ $A_2, \ldots, A_n$ are found through the recurrence relation

$$A_{k+1} = \binom{n}{k} (1-\beta)^{k} - A_k \, . \qquad (3.5.1.6)$$

The proof is a special case of theorem (T7) (5.4.2) given in chapter five, when $\lambda = 1$.

Comment

The model (3.5.1.2) though expressed for the univariate case, can be easily extended to the multivariate case. However, in some cases other time shift parameterisations are used.

For example, if instead of $\underline{J}_n(1)$ we consider $\underline{L}$, where $\underline{L}$ is an upper triangular unit matrix, then the model (3.5.1.2) can be written as

$$Y_t = \underline{f}\,\underline{\theta}_t + v_t$$

$$\underline{\theta}_t = \underline{L}\,\underline{\theta}_{t-1} + \underline{T}_t \qquad (3.5.1.7)$$

where

$v_t \sim N(0, V)$ as usual

and $\underline{T}_t \sim N(\underline{0}, \underline{H}\,\underline{W}\,\underline{H}')$

where $\underline{H}$ is a transformation matrix that transforms the dynamic system from a Jordan form to a dynamic system of an upper triangular form. The transformation matrix $\underline{H}$ is found by following the transformation and similar models procedure given in section (6.6) of chapter six.

3.5.2 E.W.R. type Polynomial D.L.M.

Extending the constant D.L.M polynomial model (3.5.1.2) to include damped polynomials so that, for $|\lambda| \leq 1$, $\underline{J}_n(\lambda)$ is the Jordan block for n equal eigenvalues , we have

$$Y_t = \underline{f}\,\underline{\theta}_t + v_t \qquad ; v_t \sim N(0, V)$$

$$\underline{\theta}_t = \underline{J}_n(\lambda)\,\underline{\theta}_{t-1} + w_t \qquad ; w_t \sim N(\underline{0}, \underline{W}) \tag{3.5.2.1}$$

where $\underline{f}$ and $\underline{\theta}_t$ have their usual meanings.

Writing the k-steps ahead forecast function

$$F_t(k) = E(Y_{t+k}\,|D_t\ ) = \underline{f}\ \{\ \underline{J}_n(\lambda)\ \}^k\ m_t \tag{3.5.2.2}$$

and $\underline{W} = \text{diag}(\ W_1, \ldots, W_n)$ where for $k=1,\ldots, n$, $0<\beta<\lambda^2$

$$W_k = \binom{n}{k}\ \{\ (\lambda^2-\beta)(1-\beta)/\beta\ \}^k\ V \tag{3.5.2.3}$$

we get in the limit $t\to\infty$ $\underline{A}_t = \underline{A} = (\ A_1, \ldots, A_n)'$

where defining $A_1 = 1 - (\beta/\lambda^2)^n$, $A_2, \ldots, A_n$ can be recursively found from the recurrence relationship

$$A_{k+1} = \binom{n}{k}\ \{\ (\lambda^2 - \beta)/\lambda\ \}^k - \lambda A_k\ . \tag{3.5.2.4}$$

For proof see theorem (T7)(5.4.2) given in chapter five.

3.5.3 The General Procedure for Deriving E.W.R. type D.L.M.

An E.W.R. type D.L.M. may be found as follows:

Using the limiting result (5.3.4.2.1), given in corollary 2 (5.3.4.2) of the G.E.W.R. theorem (5.3.4), i.e.

i) $$x_t = \prod_{i=1}^{n} (\ 1 - \lambda_i B\ ) y_t \tag{3.5.3.1}$$

and

$$x_t = \prod_{i=1}^{n} (1-\beta B/\lambda_i) e_t \qquad (3.5.3.2)$$

$$= \sum_{i=0}^{n} \mu_i e_{t-i} \qquad (3.5.3.3)$$

where the μ_i are functions of β and λ_i $(i=1,\dots,n)$.

ii) Writing the Autocovariance Generating Function (A.C.G.F.) of x_t

$$\gamma_x(B) = \prod_{i=1}^{n} \{ (1- \beta B/\lambda_i)(1- \beta B^{-1}/\lambda_i) \hat{Y} \qquad (3.5.3.4)$$

where $\hat{Y} = \lim_{t\to\infty} \hat{Y}_t = \mathrm{Var}(e_t)$.

iii) Constructing the D.L.M. in a convenient form

$$\begin{aligned} Y_t &= \underline{f}\, \underline{\theta}_t + v_t \\ \underline{\theta}_t &= \underline{G}\, \underline{\theta}_{t-1} + \underline{w}_t \end{aligned} \qquad (3.5.3.5)$$

where

$$v_t \sim N(0, V) \quad \text{and} \quad \underline{w}_t \sim N(\underline{0}, \underline{W})$$

with unknown values describing $\underline{W}$, then derive

$$x_t = f(v_t , \dots, v_{t-n} ; \underline{w}_t, \dots, w_{t-n}) \quad \text{and}$$

the A.C.G.F. $\gamma_x(B)$.

Compare with (ii) for the $\underline{W}$ setting and for $\hat{Y}$, the Var(e_t) in terms of V and β as explained in section (3.4.2)

iv) If $\lim_{t\to\infty} \underline{A}_t = \underline{A}$ is required, use

$$\begin{aligned} y_t &= \underline{f}\, \underline{G}\, \underline{m}_{t-1} + e_t , \\ \underline{m}_{t-1} &= \underline{G}\, \underline{m}_{t-2} + \underline{A}\, e_t \end{aligned} \qquad (3.5.3.6)$$

and obtain

$$x_t = \prod_{i=1}^{n} (1-\lambda_i B) \; y_t = \sum_{i=0}^{n} \mu_i \, e_{t-i} \qquad (3.5.3.7)$$

where the μ_i are linear functions of A_1 , A_2 , ... , A_n.

Comparison with (i) gives $\underline{A}$ as explained in theorem (5.4.2).

3.6 Seasonal Models

3.6.1 E.W.R. type Seasonal Models

The cases which are of particular interest are the following seasonal models, all of parsimony (2), derived when each λ_i is a root of unity and n is a period of seasonality.

i) The full pure seasonal model is

$$\sum_{i=0}^{n-1} y_{t-i} = \sum_{i=0}^{n-1} \beta^i \, e_{t-i} \, . \qquad (3.6.1.1)$$

ii) The full seasonal and level model is

$$(1-B^n) \; y_t = (1- \beta^n B^n) \; e_t \, . \qquad (3.6.1.2)$$

iii) The full linear growth seasonal model is

$$(1-B)(1-B^n) \; y_t = (1-\beta \, B)(1-\beta^n \, B^n) \; e_t \, . \qquad (3.6.1.3)$$

In general, dth order polynomial full seasonal model is

$$(1-B)^{d-1} \; (1-B^n) \; y_t = (1-\beta \, B)^{d-1} \; (1-\beta^n \, B^n) \; e_t \, . \qquad (3.6.1.4)$$

The model of $d \geq 2$ are usually not recommended for practical use.

Such models are discussed in detail by Roberts and Harrison (1982) and Roberts (1982).

3.6.2 Basic E.W.R. Cyclic Block D.L.M.

A simple cycle is characterised by a pair of distinct eigenvalues

$$e^{i\omega}\ ,\quad e^{-i\omega} \qquad ;\ \text{where}\quad \omega = 2\pi/p \quad \text{and p is the}$$

seasonal period. The model in real canonical form is defined as

$$Y_t = \underline{f}\ \underline{\theta}_t + v_t \qquad ;\quad v_t \sim N(0, V) \qquad (3.6.2.1)$$

$$\underline{\theta}_t = \underline{G}_\omega\ \underline{\theta}_{t-1} + \underline{w}_t \qquad ;\quad \underline{w}_t \sim N(0, \underline{W})$$

where

$$\underline{f} = (1 \quad 0) \quad \text{and} \quad \underline{G}_\omega = \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix}$$

The $\underline{W}$ matrix is found either following the procedure given in section (3.4.2) or using the result (3.4.3.2.8), as

$$\underline{W} = (1-\beta)\ \underline{G}_\omega \underline{C}\ \underline{G}_\omega^{'} V/\beta\ .$$

Limiting Results

The limiting forecast function for this seasonal constant D.L.M. is that derived by E.W.R. with discount factor β and the limits as $t\to\infty$ are

i) $\text{Lim Var}(e_t) = \text{Lim}\ \hat{Y}_t \to \hat{Y} = V/\beta$

ii) $\text{Lim}\ \underline{Q}_t \to \underline{Q} = \underline{f}^{'}\ \underline{f} + (\underline{G}_\omega^{'})^{-1}\ \underline{Q}\ \underline{G}_\omega^{-1}$

iii) $\text{Lim}\ \underline{A}_t \to \underline{A} = \underline{Q}^{-1}\ \underline{f} = (A_1 \quad A_2)^{'}$

where $\qquad A_1 = 1-\beta^2$

and $\qquad A_2 = (1-\beta)^2 \cot\omega$

Proof:

i) For $n=2$, writing

$$x_t = \prod_{i=1}^{2} (1-\lambda_i B) y_t = \prod_{i=1}^{2} (1-\beta B/\lambda_i) e_t$$

with $\lambda_1 = e^{i\omega}$ and $\lambda_2 = e^{-i\omega}$

we get $x_t = (1-e^{i\omega} B)(1-e^{-i\omega} B) y_t$

$= (1- \beta e^{-i\omega} B)(1-\beta e^{i\omega} B) e_t$

or $x_t = (1-2\cos \omega B+B^2) y_t$

$= (1-2 \beta\cos \omega B+\beta^2 B^2) e_t$.

The A.C.G.F. of x_t

$$\gamma_x = (1-2 \beta \cos \omega B+ \beta^2 B^2)(1-2 \beta \cos \omega B^{-1}+\beta^2 B^{-2}) \mathrm{Var}(e_t)$$

yields $\gamma_0 = (1+4\beta^2 \cos^2 \omega + \beta^4) \hat{Y}$

$\gamma_1 = -2 \beta(1+\beta^2) \cos \omega \hat{Y}$

$\gamma_2 = \beta^2 \hat{Y}$.

Now from section (3.4.2.2) we know that $\gamma_2=\lambda_1\lambda_2 V$, which in this case becomes

$\gamma_2 = V$ as $\lambda_1\lambda_2 = e^{i\omega} . e^{-i\omega} = 1$

comparing with the above we get

$$\hat{Y} = V/ \beta^2 . \qquad (3.6.2.2)$$

ii)

Writing $\underline{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix}$

and using the limiting result (2.3.3.1)

i.e $\underline{Q} = \underline{f}'\underline{f} + \beta(\underline{G}')^{-1}\underline{Q}\,\underline{G}^{-1}$ with $\underline{G} = \underline{G}_\omega$

$$\begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \beta \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix}$$

and solving these equations in terms of q's as in section (2.3.3) we get

$$q_{11} = \{(1-\beta)^{-1} + (1-\beta\cos 2\omega)/(1-2\beta\cos 2\omega+\beta^2)\}/2$$

$$q_{12} = -\{(\beta\sin 2\omega)/(1-2\beta\cos 2\omega+\beta^2)\}/2$$

$$q_{22} = \{(1-\beta)^{-1} - (1-\beta\cos 2\omega)/(1-2\beta\cos 2\omega+\beta^2)\}/2$$

Then finding the inverse of $\underline{Q}$ we get

$$\underline{Q}^{-1} = \begin{bmatrix} 1-\beta^2 & ,(1-\beta)^2\cot\omega \\ (1-\beta)^2\cot\omega, & (1-\beta)\{(3-\beta)+(1-\beta)^2\operatorname{cosec}^2\omega/\beta\} \end{bmatrix}$$

$$= \underline{C}\,V \qquad (3.6.2.3)$$

iii)

$$\underline{A} = \underline{Q}^{-1}\underline{f} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \qquad (3.6.2.4)$$

where $A_1 = 1 - \beta^2$ and $A_2 = (1-\beta)^2\cot\omega$

as required.

<u>Comment</u>:

The model (3.6.2.1) yields, in the limit, forecasts equivalent to an E.W.R. model similar to the model (3.3.2.1)

$$y_{t+i} = \underline{f}_i\underline{\theta}_t + \delta_{t+i} \quad ; \quad \delta_t \sim N(0, V)$$

with $\underline{G} = \underline{G}_\omega$

if the $\underline{W}$ matrix is set as

$$\underline{W} = (1-\beta)\, \underline{G_\omega C}\; \underline{G}_\omega' V/\beta$$

which in the present case is

$$\underline{W} = \begin{bmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{bmatrix} \qquad (3.6.2.5)$$

where

$$W_{11} = (1-\beta)(1-\beta^2)\, V/\beta^2$$

$$W_{12} = (1-\beta)^3 \cot\omega\, V/\beta^2$$

$$W_{22} = (1-\beta)\{(1-\beta^2) + (1-\beta)^3 \cot^2\omega/\beta\} V/\beta\,.$$

Another possible setting of the $\underline{W}$ matrix, given by Harrison-Akram (1983) is

$$\underline{W} = \underline{G}\;\; \underline{W}_d\;\; \underline{G}'$$

$$= \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix} \begin{bmatrix} 2\alpha & 0 \\ 0 & 2\alpha(1+\alpha/2\sin^2\omega) \end{bmatrix} x$$

$$\begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix} V$$

$$= \alpha \begin{bmatrix} 2+\alpha & \alpha\cos\omega \\ \alpha\cos\omega & 2+\alpha\cot^2\omega \end{bmatrix} V \qquad (3.6.2.6)$$

where $\underline{W}_d = \mathrm{diag}(\ W_1\ ,\ W_2\ )$

$$W_1 = 2\,\alpha \qquad \text{and} \qquad W_2 = 2\,\alpha\,(\ 1 + \alpha\,/\ 2\sin^2 \omega\ )$$

and $\qquad \alpha = (\ 1 - \beta\ )^2\ /\ \beta$

3.6.3 An E.W.R. Seasonal Constant D.L.M.

This model is a full seasonal and level (of a stochastic process) model with a stochastic 'form free' seasonal pattern, varying about a stochastic process level.

Let the seasonal cycle be of n (integer) sampling intervals, then the eigenvalues are

$$e^{2\pi k\, i/n} \qquad ;\ k = 0,1,\ldots\ldots,n-1$$

and the model can be written in a form suitable for operation as (see Harrison-Akram(1983))

$$y_t = \underline{f}\ \underline{\theta}_t + v_t$$

$$\underline{\theta}_t = \begin{bmatrix} \underline{0} & \underline{I}_{n-1} \\ 1 & \underline{0}' \end{bmatrix} \underline{\theta}_{t-1} + \underline{w}_t \qquad (3.6.3.1)$$

where $\quad v_t \sim N(\ 0,\ V\ ) \quad$ and $\quad \underline{w}_t \sim N(\ \underline{0},\ \underline{W}\ )$

with as usual $\quad \underline{f} = (1,0,\ldots\ldots,0)$.

Writing $\quad W = (\ 1 - \beta^n\ )^2\ V/\ n\,\beta^n \qquad (3.6.3.2)$

and setting $\underline{W} = W\ \underline{I}$, the model is such that

i) the limiting forecast function is that derived for a full seasonal and level D.L.M. using E.W.R. with discount factor β

ii) $\lim_{t\to\infty} \mathrm{Var}\ (\ Y_t\ |\ D_t\ ) = V/\ \beta^n$

iii) $\lim_{t \to \infty} \underline{A}_t = \underline{A} = (A_1, \ldots, A_n)'$ where $A_1 = 1-\beta^n$, $A_i = 0$ for all $i \geq 2$.

<u>Proof</u>

From the definition of the full seasonal and level model (3.6.1.2)

$$x_t = (1-B^n)y_t \tag{3.6.3.3}$$

$$= (1-\beta^n B^n)e_t. \tag{3.6.3.4}$$

Now from the definition of the model (3.6.3.1), writing the observation y_t at time t as

$$y_t = \theta_{1,t} + v_t \tag{3.6.3.5}$$

$$= (\theta_{2,t-1} + w_{1,t}) + v_t$$

$$\vdots$$

$$= \theta_{n,t+1-n} + \sum_{i=1}^{n-1} w_{i,t+1-i} + v_t$$

$$= B^n \theta_{1,t} + \sum_{i=1}^{n} w_{i,t+1-i} + v_t. \tag{3.6.3.6}$$

This shows that

$$\theta_{1,t} = B^n\, \theta_{1,t} + \sum_{i=1}^{n} w_{i,t+1-i} \text{ or } (1-B^n)\, \theta_{1,t} = \sum_{i=1}^{n} w_{i,t+1-i} \tag{3.6.3.7}$$

Multiplying both sides of (3.6.3.5) by $(1-B^n)$ and substituting this result we get

$$(1-B^n)y_t = \sum_{i=1}^{n} w_{i,t+1-i} + (1-B^n)v_t \tag{3.6.3.8}$$

Equating for the autocovariances of x_t

$$\gamma_k = \begin{cases} (1+\beta^{2n})V/\beta^n & \text{for } k=0 \\ -\beta^n\, V(e) & \text{for } k=n \\ 0 & \text{otherwise} \end{cases} \tag{3.6.3.9}$$

gives confirmation of (i) and (ii).

We know that an observation

$$y_t = m_{2,t-1} + e_t \tag{3.6.3.10}$$

can be written from the forecast function.

Now

$$\underline{m}_t = \begin{bmatrix} \underline{0} & \underline{I}_{n-1} \\ 1 & \underline{0}' \end{bmatrix} \underline{m}_{t-1} + \underline{A}\, e_t$$

as $t \to \infty$. From it finding the value of $m_{2,t-1}$ in a recursive manner and substituting in (3.6.3.3) we get

$$(1-B^n)\, y_t = e_t + \sum_{i=2}^{n} A_i\, e_{t+1-i} - (1-A_1)\, e_{t-n}$$

$$= (1- \beta^n B^n)\, e_t \tag{3.6.3.11}$$

which confirms (iii).

<u>Comment</u>

For some models a seasonal component is required in which the parameters are seasonal effects. It is ofcourse easy to ensure that the sum of parameters is fixed, say at zero when they then represent such effects.

Initially set $\underline{m}_0 = \underline{0}$ and set the associated variance $\underline{C}_0$ so that all rows and columns sum to zero.

The additional constraint is such that the elements of $\underline{w}_t$ sum to zero over a seasonal period n . This is due to the assumption that the seasonality is normalised over a cycle or period n such that

$$\sum_{i=1}^{n} s_{i,t} = 0 = \sum_{i=1}^{n} w_{i,t}$$

One way of ensuring that these constraints are met is as follows:

Let the prior information be expressed as

$$(\underline{w}_t \mid D_{t-1}) \sim N(\hat{\underline{w}}_t , \underline{W}_n).$$

Suppose that the constraints are not all satisfied and the joint prior of

$$\underline{w}_t \quad \text{and} \quad \sum_{i=i}^{n} w_{i,t}$$

is

$$\begin{pmatrix} \underline{w}_t \\ \sum_{i=1}^{n} w_{i,t} \end{pmatrix} \sim N \begin{bmatrix} \hat{\underline{w}}_t & \\ \sum_{i=1}^{n} \hat{w}_{i,t} & , \quad \underline{W}_{n+1} \end{bmatrix} \tag{3.6.3.12}$$

where

$$\underline{W}_{n+1} = \begin{bmatrix} \underline{W}_n & \underline{W} \\ \underline{W}' & W_s \end{bmatrix}$$

$$\underline{W} = (W_1 , \ldots , W_n)'.$$

W_i is the sum of the elements in the i-th row of $\underline{W}_n$ and W_s is the sum of all elements in $\underline{W}_n$.

Defining $\underline{A} = \underline{W} / W_s$

for $\sum_{i=1}^{n} w_{i,t} = 0$

$$\underline{w}_t \sim N (\hat{\underline{w}}_t^* , \underline{W}_n^*) \tag{3.6.3.13}$$

where

$$\hat{\underline{w}}_t^* = \hat{\underline{w}}_t - \underline{A} \sum_{i=1}^{n} \hat{w}_{i,t}$$

$$\underline{W}_n^* = \underline{W}_n - \underline{A}\,\underline{A}'\, W_s$$

Writing $\underline{W}_n = \{ W_{ij} \}$, we get

$$\underline{W}_n^* = \{ W_{ij} - W_i W_j / W_s \} = \{ W_{ij}^* \} . \qquad (3.6.3.14)$$

As a special case when $\underline{W}_n = W \underline{I}_n$, this result reduces to

$$W_{ij}^* = \begin{cases} (n-1) W / n & ; \text{ if } i = j \\ - W/n & ; \quad i \neq j \end{cases}$$

A result given by Harrison-Akram (1983).

Such constrained models are in use in both seasonal and non seasonal Bayesian forecasting applications. In addition, in practice we do not demand that the model be constant but generally keep the ratio V/W constant. Hence the uncertain-ities which may well depend upon the expected level of the observation (particularly important in seasonal cases) can be more properly described, while retaining the concept of an E.W.R. forecasting function.

4

RECURRENCE RELATIONSHIPS FOR THE PRECISION MATRICES OF AUTOREGRESSIVE - MOVING AVERAGE PROCESSES

4.1 Introduction

The introduction of Autoregressive processes by Yule (1927) and Moving Average processes by Slutzky (1927), has generated much interest. The work in this area has steadily been advanced by Walker (1931), Wold (1938), Bartlet (1946), Kendall (1949), Durbin (1959), Hannan (1969), Akaike (1971), Box-Jenkins (1974,76), Godolphin (1977,80), Nuri (1980), etc.

In this chapter I discuss coloured noise processes which display Autoregressive-Moving Average (ARMA) behaviour. Recurrence relations which play a dominent role in dynamic systems are developed for ARMA precision matrices. The attention is, however, restricted to those precision matrices which are relevant to ARMA(p,q) processes.

Special cases such as pure Autoregressive processes and pure Moving Average processes together some examples are discussed.

In the last section, a brief survey of the methodology of Box-Jenkins (1976) relating to ARMA processes is given.

4.2 (p,q) Order Autoregressive - Movivg Average (ARMA(p,q)) Coloured Noise Process

For an ARMA(p,q) coloured noise process ε_t with representation

$$\phi_p(B)\ \varepsilon_t = \eta_q(B)\ \delta_t \tag{4.2.0.1}$$

where δ_t is a white noise with variance σ_δ^2 ,

$$\phi_p(B) = \prod_{i=1}^{p} (\ 1 - \phi_i\ B\ )$$

and
$$\eta_q(B) = \prod_{i=1}^{q} (\ 1 - \eta_i\ B\ )$$

are polynomials in B of degrees p and q respectively, the roots of which are assumed to lie outside the unit circle; defining:

i) $\psi(B) = \phi_p(B)/\ \eta_q(B)$

$$= \sum_{i=0}^{\infty} \psi_i\ B^i$$

where $\psi_0 = 1$

ii) $$\psi_t(B) = \sum_{i=0}^{t-1} \psi_i \, B^i$$

iii) (1xt) vectors

$$\underline{\psi}_t = (1, \psi_1, \ldots, \psi_{t-1})'$$

$$\underline{\delta}_t = (\delta_t, \ldots, \delta_1)'$$

$$\underline{\varepsilon}_t = (\varepsilon_t, \ldots, \varepsilon_1)'$$

where $\underline{\varepsilon}_t \sim (\underline{0}, \underline{P}_t^{-1} \sigma_\delta^2)$;

and assuming that $\varepsilon_t = 0$ for $t \leq 0$ and the inverse of $\underline{P}_t$, the precision matrix, i.e. $\underline{P}_t^{-1}$ exists; the recurrence relation for $\underline{P}_t$ is

$$\underline{P}_t = \underline{\psi}_t \ \underline{\psi}_t' + \begin{bmatrix} 0 & \underline{0}_{t-1}' \\ \underline{0}_{t-1} & \underline{P}_{t-1} \end{bmatrix}$$

with $\underline{0}_{t-1}'$ as the (t-1) row vector of zeros and $P_1 = 1$.

Proof:

Since $\phi_p(B)$ and $\eta_q(B)$ are invertible, initialisation does not effect limiting results and generally the effect dies away quickly with time. Hence, considering assumption

$$\varepsilon_t = 0 \qquad \text{for} \quad t \leq 0$$

we can write

$$\delta_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots + \psi_{t-1}\varepsilon_1$$

$$\delta_{t-1} = \varepsilon_{t-1} + \psi_1\varepsilon_{t-2} + \cdots + \psi_{t-2}\varepsilon_1$$

$$\vdots$$

$$\delta_1 = \varepsilon_1$$

or $$\underline{\delta}_t = \underline{M}_t \, \underline{\varepsilon}_t \tag{4.2.0.2}$$

where

$$\underline{M}_t = \begin{pmatrix} \underline{\psi}'_t \\ \\ \underline{N}_t \end{pmatrix} \quad ; \tag{4.2.0.3}$$

$$\underline{N}_t = (\underline{0}_{t-1} \quad \underline{M}_{t-1}) \, . \tag{4.2.0.4}$$

From equation (4.2.0.1) we can write

$$\underline{\varepsilon}_t = \underline{M}_t^{-1} \, \underline{\delta}_t \tag{4.2.0.5}$$

$$\mathrm{Var}(\underline{\varepsilon}_t) = \underline{M}_t^{-1} \, (\underline{M}_t')^{-1} \, \sigma_\delta^2$$

$$= \underline{P}_t^{-1} \, \sigma_\delta^2 \tag{4.2.0.6}$$

where

$$\underline{P}_t^{-1} = \underline{M}_t^{-1} \, (\underline{M}_t')^{-1} \, . \tag{4.2.0.7}$$

From this expression of $\underline{P}_t^{-1}$ we can easily write

$$\underline{P}_t = \underline{M}_t' \, \underline{M}_t$$

$$= \begin{pmatrix} \underline{\psi}_t & \underline{N}_t' \end{pmatrix} \begin{pmatrix} \underline{\psi}_t' \\ \underline{N}_t \end{pmatrix}$$

$$= \underline{\psi}_t \, \underline{\psi}_t' + \underline{N}_t' \, \underline{N}_t$$

Substituting the value of $\underline{N}_t$ from the expression (4.2.0.4) here, we get

$$\underline{P}_t = \underline{\psi}_t \; \underline{\psi}_t' + \begin{pmatrix} \underline{0}_{t-1}' \\ \underline{M}_{t-1}' \end{pmatrix} (\underline{0}_{t-1} \quad \underline{M}_{t-1})$$

$$= \underline{\psi}_t \; \underline{\psi}_t' + \begin{bmatrix} 0 & \underline{0}_{t-1}' \\ \underline{0}_{t-1} & \underline{P}_{t-1} \end{bmatrix} \qquad (4.2.0.8)$$

where $\underline{P}_1 = 1$.

<u>Example</u>

Let $\varepsilon_t \sim$ ARMA(1,1) with representation

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \delta_t - \eta_1 \delta_{t-1}$$

then at time t=3 if we wish to find the precision matrix $\underline{P}_3$ of the given process ε_t, it can be evaluated easily through the above recurrence relation (4.2.0.8) as follows.

At time t = 1, we know that

$P_1 = 1$.

Substituting this value of P_1 in to the recurrence relation of $\underline{P}_2$, i.e.

$$\underline{P}_2 = \underline{\psi}_2 \ \underline{\psi}_2' + \begin{bmatrix} 0 & 0 \\ 0 & P_1 \end{bmatrix}$$

$$= \begin{pmatrix} 1 \\ \psi_1 \end{pmatrix} (1 \quad \psi_1) + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \psi_1 \\ \psi_1 & 1 + \psi_1^2 \end{bmatrix}$$

Substituting this result further in to the recurrence relation of $\underline{P}_3$, i.e.

$$\underline{P}_3 = \underline{\psi}_3 \ \underline{\psi}_3' + \begin{bmatrix} 0 & \underline{0}_2' \\ \underline{0}_2 & \underline{P}_2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \psi_1 & \psi_2 \\ \psi_1 & 1+\psi_1^2 & \psi_1(1+\psi_2) \\ \psi_2 & \psi_1(1+\psi_2) & 1+\psi_1^2+\psi_2^2 \end{bmatrix}$$

4.3 Special Cases

4.3.1 p-th Order Autoregressive (AR(p)) Process

This is a special case of ARMA(p,q) process (4.2.0.1) when

$$\eta_i = 0 \quad \text{for all} \quad 1 \leq \eta_i \leq q.$$

In this case the coloured noise $\varepsilon_t \sim$ ARMA(p,0) or simply AR(p) with representation

$$\phi_p(B)\,\varepsilon_t = \delta_t \tag{4.3.1.1}$$

The recurrence relation (4.2.0.8), i.e.

$$\underline{P}_t = \underline{\psi}_t\ \underline{\psi}_t' + \begin{bmatrix} 0 & \underline{0}_{t-1}' \\ \underline{0}_{t-1} & \underline{P}_{t-1} \end{bmatrix}$$

holds with

$$\psi(B) = \phi_p(B) = \prod_{i=1}^{p} (1 - \phi_i B)$$

$$= \sum_{i=0}^{p} \psi_i B^i \tag{4.3.1.2}$$

and $\quad P_1 = 1$ as usual.

Further, if $p = 1$, (4.3.1.1) reduces to a first order Autoregressive (AR(1)) process

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \delta_t . \qquad (4.3.1.3)$$

This is also known as Markov process.

In this case

$$\underline{\psi}_t = \underline{\phi}_t = (1, - \phi_1, 0, \ldots, 0)'$$

$$\psi(B) = \phi_1(B) = 1 - \phi_1 B$$

and the recurrence relation (4.2.0.8) reduces to

$$\underline{P}_t = \underline{\phi}_t \underline{\phi}_t' + \begin{bmatrix} 0 & \underline{0}_{t-1}' \\ \underline{0}_{t-1} & \underline{P}_{t-1} \end{bmatrix} \qquad (4.3.1.4)$$

<u>Example</u>

Let $\varepsilon_t \sim$ ARMA(1,0), i.e. AR(1) with representation (4.3.1.3), then at time $t=3$, the precision matrix $\underline{P}_3$ can be evaluated recursively through the above recurrence relation (4.3.1.4) as

$$P_1 = 1$$

$$\underline{P}_2 = \begin{pmatrix} 1 \\ -\phi_1 \end{pmatrix} (1 \quad -\phi_1) + \begin{bmatrix} 0 & 0 \\ 0 & P_1 \end{bmatrix}$$

or $$\underline{P}_2 = \begin{bmatrix} 1 & -\phi_1 \\ -\phi_1 & 1+\phi_1^2 \end{bmatrix}$$

Substituting this value of $\underline{P}_2$ in to the recurrence relation of $\underline{P}_3$, i.e.

$$\underline{P}_3 = \underline{\phi}_2 \, \underline{\phi}_2' + \begin{bmatrix} 0 & \underline{0}_2' \\ \underline{0}_2 & \underline{P}_2 \end{bmatrix}$$

where $\underline{\phi}_2 = (1 \quad -\phi_1 \quad 0)'$

we get

$$\underline{P}_3 = \begin{bmatrix} 1 & -\phi_1 & 0 \\ -\phi_1 & 1+\phi_1^2 & -\phi_1 \\ 0 & -\phi_1 & 1+\phi_1^2 \end{bmatrix}$$

the required precision matrix at time $t = 3$.

4.3.2 q-th Order Moving Average (MA(q)) Process

This is a special case of (4.2.0.1) when the Autoregressive coefficients

$$\phi_i = 0 \quad \text{for all} \quad 1 \leq i \leq p.$$

In this case the coloured noise $\varepsilon_t \sim$ ARMA(0,q) or MA(q) with representation

$$\varepsilon_t = \eta_q(B)\ \delta_t \tag{4.3.2.1}$$

The recurrence relation (4.2.0.8) for the precision matrix $\underline{P}_t$, i.e.

$$\underline{P}_t = \underline{\psi}_t\ \underline{\psi}_t' + \begin{bmatrix} 0 & \underline{0}_{t-1}' \\ \underline{0}_{t-1} & \underline{P}_{t-1} \end{bmatrix}$$

holds with

$$(B) = 1/\eta_q(B) = \sum_{i=0}^{\infty} \psi_i\ B^i \tag{4.3.2.2}$$

and $\quad P_1 = 1 \quad$ as usual.

In particular, if q = 1 then $\varepsilon_t \sim$ MA(1) with representation

$$\varepsilon_t = \delta_t - \eta_1 \delta_{t-1} \tag{4.3.2.3}$$

For this first order Moving Average process

$$\underline{\psi}_t = \underline{\eta}_t = (\ 1,\ \eta_1,\ \eta_1^2, \ldots,\ \eta_1^{t-1}\ )' \tag{4.3.2.4}$$

and

$$\underline{P}_t = \underline{n}_t \; \underline{n}_t' + \begin{bmatrix} 0 & \underline{0}_{t-1}' \\ \underline{0}_{t-1} & \underline{P}_{t-1} \end{bmatrix} \qquad (4.3.2.5)$$

<u>Example</u>

For a coloured noise process $\varepsilon_t \sim$ MA(1) with representation (4.3.2.3), the precision matrix of ε_t at time $t = 3$, i.e. $\underline{P}_3$ can be evaluated through the above recurrence relation (4.3.2.5) as follows.

We know that at time $t = 1$ $P_1 = 1$. Substituting this value of P_1 in to the recurrence relation of $\underline{P}_2$, i.e.

$$\underline{P}_2 = \begin{pmatrix} 1 \\ n_1 \end{pmatrix} (1 \quad n_1) + \begin{bmatrix} 0 & 0 \\ 0 & P_1 \end{bmatrix}$$

we get

$$\underline{P}_2 = \begin{bmatrix} 1 & n_1 \\ n_1 & 1 + n_1^2 \end{bmatrix}$$

Further substituting this value of $\underline{P}_2$ in to the recurrence relation for the precision matrix $\underline{P}_3$, i.e.

$$\underline{P}_3 = \begin{bmatrix} 1 \\ \eta_1 \\ \eta_1^2 \end{bmatrix} (1 \quad \eta_1 \quad \eta_1^2) + \begin{bmatrix} 0 & \underline{0}'_{t-1} \\ \underline{0}_{t-1} & \underline{P}_2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \eta_1 & \eta_1^2 \\ \eta_1 & 1 + \eta_1^2 & \eta_1(1 + \eta_1^2) \\ \eta_1^2 & (1 + \eta_1^2) & 1 + \eta_1^2 \quad \eta_1^4 \end{bmatrix} \tag{4.3.2.6}$$

<u>Comment</u>:

The recurrence relations, developed for ARMA(p,q) precision matrices play quite a significant role in the development of Generalised Exponentially Weighted Regression (G.E.W.R.) theory, presented in the forthcoming chapters.

In case of Exponentially Weighted Regression (E.W.R.) the precision matrix $\underline{P}_t$ is simply equal to $\underline{I}_t$, i.e. an identity matrix at time t.

4.4 ARMA Processes and Box Jenkins Models

Box and Jenkins in a series of papers and a subsequent book (1976), introduced a set of univariate models. they presented ARMA Models for stationary time series and ARIMA Models for non-stationary time series, assuming that the non-stationarity can be transformed to stationarity by virtue of differencing the series up to an appropriate order $d \geq 1$. The models introduced can be adopted to take in to account seasonal variation. They consider the models with time invariant parameters and proclaim the parsimony. Their approach to time series analysis is characterised by a very sophisticated three stage procedure for model identification, parameter estimation and forecasting. The detail can be found in Box-Jenkins (1976).

During the last decade their methodology has obtained quite extensive publicity, perhaps due to a claim by Box and Jenkins that their ARIMA Models are almost universal, covering almost all previous model conceptions such as decomposition of time series in to trend plus seasonality plus stationary fluctuations; but the pasture is not so green as promised. The methodology suffers from many drawbacks, such as:

1) ARIMA Models do not provide an explanatory insight in to the trend, seasonality and stationary residual components of a time series. When the variation of the systematic part such as trend and seasonality is dominant, the prediction effectiveness of the ARIMA Model is mainly determined by the initial differncing operations

$$\nabla^d = (1 - B)^d \quad \text{and} \quad \nabla_s^D = (1 - B^s)^D$$

where d and D are the degrees of differing for trend and seasonality and s is the seasonal period, and not by the ARMA Model fitted to the stationary part. For more discussion see Parzen (1974). The differencing, usually creats a problem, especially, when the time series contain outliers. In such a case the differencing increase the number of outliers.

For example, the first differences produce two outliers (in doublet form) for every isolated outlier in original series. In general the outliers are moutiplied by a factor ζ , such that

$$\zeta \cong 2(d + D)$$

2) Due to the non-adaptive nature of the parameter system in Box-Jenkin methodology, the ARIMA Models

i) are inflexible as compared to the Dynamic Linear Models system

ii) are incapable of handling abrupt structural changes in the time series

iii) do not provide automatic procedures for updating the estimates of model parameters with the arrival of new data

3) Box-Jenkins models are not easy to identify without a great deal of practical knowledge and expertise.

4) By contrast to the sophistication offered by the Box-Jenkins methodology, other simple and straightforward methods such as EXponentially Weighted Regression (E.W.R.)

methods quite often provide reasonable estimates. For related discussion see Wilson and Newman (1981).

The Generalised Exponentially Weighted Regression (G.E.W.R.) methodology, presented in the forthcoming chapters, overcomes most of the stated problems.

5

GENERALISED EXPONENTIALLY WEIGHTED REGRESSION

G . E . W . R

5.1 Introduction

The infrastructure developed in the previous chapters is used in this chapter to develop a dynamic Bayesian methodology for the analysis and forecasting of discrete time series with ARMA type Coloured Noise processes.

The aim of this generalization is to retain the conceptual simplicity and parsimony of E.W.R. for longer term, trend or low frequency movements and to model the higher frequencies separately as ARMA type Coloured Noise processes.

In section two Generalised Exponentially Weighted Regression (G.E.W.R.) is defined. Some basic components, involved in the development of the dynamic system of the G.E.W.R. are explained in detail.

In section three, firstly, a Fundamental Theorem of G.E.W.R. is given along with limiting results; then the Main Theorem of G.E.W.R. is presented, which apart from generalizing the results of Godolphin-Harrison (1975) and Mckenzie (1976), provides a technically concise proof, which is shorter than that used for the special cases. The rest of the section is devoted to the development of the G.E.W.R. theory through various theorems. Many corollaries to the theorems are presented.

In section four, some practical implications of G.E.W.R. are discussed. Two useful theorems, in this connection are given.

5.2 Definitions

5.2.1 Definition of G.E.W.R.

For a time sequence Y_t , consider the model at time t as

$$Y_{t-i} = \underline{f}_{t-i}\underline{\theta} + \varepsilon_{t-i} \tag{5.2.1.1}$$

where the error $\underline{\varepsilon}_t' = (\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_1)$ is such that

$$\underline{\varepsilon}_t \sim (\underline{0} ; \underline{P}_t^{-1} \sigma_\delta^2)$$

and the independent variable 1xn vectors $\underline{f}_i$ are known, and for smoothing or discount factor $0 < \beta < 1$, let $\underline{\beta}_t^{1/2}$ be a t-diagonal matrix such that

$$\underline{\beta}_t^{1/2} = \text{diag}\,(1, \beta^{1/2}, \ldots, \beta^{(t-1)/2}),$$

then the G.E.W.R. estimate $\underline{m}_t$ for θ based on observations $(y_t, y_{t-1}, \ldots, y_1)$ is defined as that vector value of $\underline{\theta} \in R^n$ which minimizes

$$R_t = \underline{\varepsilon}_t' \underline{\beta}_t^{1/2} \underline{P}_t \underline{\beta}_t^{1/2} \underline{\varepsilon}_t \tag{5.2.1.2}$$

or equivalently maximizes

$$S_t = - R_t/2 \tag{5.2.1.3}$$

Clearly if $\underline{P}_t = \underline{I}_t$ then $\underline{m}_t$ is the usual E.W.R. estimate and if the joint distribution of $\underline{\varepsilon}_t$ is normal, $\underline{m}_t$ is the Maximum Discounted Likelihood estimate.

5.2.2 <u>Definition of the Components of G.E.W.R.</u>

$$\psi_t(B) = \sum_{i=0}^{t-1} \psi_i B^i \tag{5.2.2.1}$$

$$\alpha_i = \psi_i \beta^{i/2} \quad \text{for } i \geq 0 \tag{5.2.2.2}$$

$$\underline{\alpha}_t = (\alpha_0, \alpha_1, \ldots \alpha_{t-1})' = \underline{\beta}_t^{1/2} \underline{\psi}_t \tag{5.2.2.3}$$

$$\underline{\psi}_t \quad (\psi_0, \psi_1, \ldots \psi_{t-1})' \tag{5.2.2.4}$$

$$\alpha_t(B) = \psi_t(\beta^{1/2} B) = \sum_{i=0}^{t-1} \alpha_i B^i \qquad (5.2.2.5)$$

$$\underline{F}_t' = (\underline{f}_{t-1}', \underline{f}_{t-2}', \ldots \underline{f}_0') \qquad (5.2.2.6)$$

$$\underline{y}_t' = (y_t, y_{t-1}, \ldots y_1), \text{ original series} \qquad (5.2.2.7)$$

$$\underline{u}_t = \underline{\alpha}_t' \underline{F}_t = \sum_{i=0}^{t-1} \alpha_i \underline{f}_{t-1-i} = \sum_{i=0}^{t-1} \psi_i \beta^{i/2} \underline{f}_{t-1-i} \qquad (5.2.2.8)$$

$$z_t = \underline{\alpha}_t' \underline{y}_t = \psi_t(\beta^{1/2} B) y_t, \text{ derived series.} \qquad (5.2.2.9)$$

$$d_t = \underline{u}_t \underline{m}_{t-1}, \text{ one step ahead forecast of the derived series.} \qquad (5.2.2.10)$$

$$e_t = z_t - d_t.$$

Example

Some of the definitions, given in previous section and above, may be illustrated as follows.

For $\varepsilon_t \sim \text{ARMA}(p,q)$, given discount factor β and known $\underline{f}_t$, we see that:

i) at time $t = 1$, $\underline{\beta}_1^{1/2} = 1$, $\underline{P}_1 = 1$, $\underline{\varepsilon}_1 = \varepsilon_1$ and $\underline{\alpha}_1 = 1$, which gives $S_1 = -\varepsilon^2/2$, $\underline{u}_1 = \underline{f}_0$ and $z_1 = y_1$;

ii) at time $t = 2$, $\underline{\beta}_2^{1/2} = \text{diag}(1, \beta^{1/2})$, $\underline{P}_2 = \begin{bmatrix} 1 & \psi_1 \\ \psi_1 & 1+\psi_1^2 \end{bmatrix}$

$\underline{\varepsilon}_2' = (\varepsilon_2 \;\; \varepsilon_1)$ and $\underline{\alpha}_2 = (\alpha_0 \;\; \alpha_1)' = (1 \;\; \psi_1\beta^{1/2})'$; which gives

$$S_2 = -(\varepsilon_2 + \psi_1\beta^{1/2}\varepsilon_1)^2/2 + \beta S_1 ,$$

$$\underline{u}_2 = \underline{f}_1 + \psi_1\beta^{1/2}\underline{f}_0 \quad \text{and} \quad z_2 = y_2 + \psi_1\beta^{1/2} y_1.$$

and so on. More detailed description and applications of these definitions are given in various forthcoming sections of the thesis.

5.2.3 System Matrix $\underline{Q}_t$

The system matrix $\underline{Q}_t$ of G.E.W.R. is defined as

$$\underline{Q}_t = \sum_{i=0}^{t-1} \beta^i \underline{u}_{t-i}' \underline{u}_{t-i} = \underline{u}_t'\underline{u}_t + \sum_{i=1}^{t-1} \beta^i \underline{u}_{t-i}' \underline{u}_{t-i} \qquad (5.2.3.1.1)$$

Re-arranging the terms, we get

$$\underline{Q}_t = \underline{u}_t' \underline{u}_t + \beta \sum_{i=0}^{t-2} \beta^i \underline{u}_{t-1-i}' \underline{u}_{t-1-i} \qquad (5.2.3.1.2)$$

$$= \underline{u}_t' \underline{u}_t + \beta \underline{Q}_{t-1} \qquad (5.2.3.1.3)$$

where the matrix $\underline{Q}_t$ is of full rank n.

5.2.3.2 Time Series Case

For, time series, the usual moving origin representation is adopted so that at time t,

$$Y_{t+i} = \underline{f}_i \underline{\theta} + \varepsilon_{t+i} \qquad (5.2.3.2.1)$$

where $\underline{f}_i = \underline{f}_0 \underline{G}^i$, $(\underline{f}_0', \underline{f}_1', \ldots, \underline{f}_{n-1}')$ is of rank n and $\underline{G}$, the transition matrix, has non zero eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ which are not necessarily distinct.

In this case the expression for the $\underline{Q}_t$ matrix is

$$\underline{Q}_t = \underline{u}_t' \underline{u}_t + \beta (\underline{G}')^{-1} \underline{Q}_{t-1} \underline{G}^{-1} . \qquad (5.2.3.2.2)$$

When $\underline{G} = \underline{I}$, the identity matrix, the result obviously reduces to the expression (5.2.3.1.3).

If $\psi_t(B) = \phi_t(B) = \eta_t(B) = 1$, then $\underline{P}_t = \underline{I}_t$.
In this case $\underline{u}_t = \underline{f}_t$ and the expression (5.2.3.2.2) reduces to (2.3.1.4), i.e. the expression for $\underline{Q}_t$ in the E.W.R. and D.E.W.R. case.

The inverse of $\underline{Q}_t$, if required, can be found by using the vector-matrix mixed identity (3.2.2.1).

5.2.3.3 Limiting Form

In the limit as $t \to \infty$

$$\underline{u}_t \to \underline{u} = \sum_{i=0}^{\infty} \alpha_i \underline{f}_i$$

$$\underline{Q}_t \to \underline{Q} = \underline{u}' \underline{u} + \beta (\underline{G}')^{-1} \underline{Q} \underline{G}^{-1} \qquad (5.2.3.3.1)$$

5.2.4 System Vector $\underline{H}_t$

5.2.4.1 Ordinary Case

The system vector $\underline{H}_t$ is defined as

$$\underline{H}_t = \sum_{i=0}^{t-1} (\beta B)^i \underline{u}_t' z_t$$

$$= \sum_{i=0}^{t-1} \beta^i \underline{u}_{t-i}' z_{t-i} \qquad (5.2.4.1.1)$$

$$= \underline{u}_t' z_t + \beta \sum_{i=1}^{t-1} \beta^{i-1} \underline{u}_{t-i}' z_{t-i}$$

Re-arranging the terms we get

$$\underline{H}_t = \underline{u}_t' z_t + \beta \sum_{i=0}^{t-2} \beta^i \underline{u}_{t-1-i}' z_{t-1-i}$$

$$= \underline{u}_t' z_t + \beta \underline{H}_{t-1} . \qquad (5.2.4.1.2)$$

5.2.4.2 Time Series Case

Considering the usual moving origin representation, the system vector $\underline{H}_t$ is defined as

$$\underline{H}_t = \sum_{i=0}^{t-1} \{ \beta(\underline{G}')^{-1} B \}^i \underline{u}_t' z_t \qquad (5.2.4.2.1)$$

$$= \underline{u}_t' z_t + \beta(\underline{G}')^{-1} \underline{H}_{t-1} . \qquad (5.2.4.2.2)$$

5.2.4.3 Limiting Form

In the limit $t \to \infty$, $\underline{u}_t \to \underline{u}$

and

$$\underline{H}_t = \underline{u}' z_t + \beta(\underline{G}')^{-1} \underline{H}_{t-1} . \qquad (5.2.4.3.1)$$

5.3 Theorems Of G.E.W.R.

5.3.1 Fundamental Theorem (T2)(Ordinary Case)

Following the definitions given in section (5.2.2), the G.E.W.R. estimate $\underline{m}_t$ for a parameter vector $\underline{\theta}$ is such that

$$\underline{m}_t = \underline{Q}_t^{-1} \underline{H}_t$$

$$= \underline{m}_{t-1} + \underline{A}_t e_t \qquad (5.3.1.1)$$

where $\underline{A}_t = \underline{Q}_t^{-1} \underline{u}_t'$.

Proof:

For the model

$$Y_{t-i} = \underline{f}_{t-i} \underline{\theta} + \varepsilon_{t-i} \qquad (5.3.1.2)$$

following the definitions given in section (5.2.1) and the recurrence relations for the precision matrices (4.3.0.2) and defining $S_0 = 0$, we can write (5.2.1.3) at time t, recursively as

$$S_t = \beta S_{t-1} - \underline{\varepsilon}_t' \underline{a}_t \underline{a}_t' \underline{\varepsilon}_t /2 \qquad (5.3.1.3)$$

For $t > 1$ we can write the expression (5.3.1.3) as

$$S_t = \beta S_{t-1} - (z_t - \underline{u}_t \underline{\theta})^2/2$$

and for $t = 1$

$$S_1 = - (z_1 - \underline{u}_1 \underline{\theta})^2/2 . \qquad (5.3.1.4)$$

Hence $$S_t = - \sum_{i=0}^{t-1} \beta^i (z_{t-i} - \underline{u}_{t-i} \underline{\theta})^2/2 \qquad (5.3.1.5)$$

which is minimised when $\underline{\theta} = \underline{m}_t$ where

$$\left.\frac{dS_t}{d\underline{\theta}}\right|_{\underline{m}_t} = \underline{0} = \left.\sum_{i=0}^{t-1} \beta^i \underline{u}'_{t-i} (z_{t-i} - \underline{u}_{t-i} \underline{\theta})\right|_{\underline{m}_t}$$

giving $$\sum_{i=0}^{t-1} \beta^i \underline{u}'_{t-i} \underline{u}_{t-i} \underline{m}_t = \sum_{i=0}^{t-1} \beta^i \underline{u}'_{t-i} z_{t-i} . \qquad (5.3.1.6)$$

Now from definitions (5.2.3.1.1) and (5.2.4.1.1) we know that

$$\sum_{i=0}^{t-1} \beta^i \underline{u}'_{t-i} \underline{u}_{t-i} = \underline{Q}_t \quad \text{and} \quad \sum_{i=0}^{t-1} \beta^i \underline{u}'_{t-i} z_{t-i} = \underline{H}_t .$$

Substituting these into (5.3.1.6) we get

$$\underline{Q}_t \underline{m}_t = \underline{H}_t \qquad (5.3.1.7)$$

or $$\underline{m}_t = \underline{Q}_t^{-1} \underline{H}_t . \qquad (5.3.1.8)$$

Multiplying both sides of the expression (5.3.1.7) by βB we get

$$\beta \underline{Q}_{t-1} \underline{m}_{t-1} = \beta \underline{H}_{t-1} . \qquad (5.3.1.9)$$

Now from (5.2.4.1.2)

$$\underline{H}_t = \underline{u}'_t z_t + \beta \underline{H}_{t-1}$$

so (5.3.1.7) becomes

$$\underline{Q}_t \, \underline{m}_t = \underline{u}_t' \, z_t + \beta \, \underline{H}_{t-1}$$

$$= \underline{u}_t' \, z_t + \beta \, \underline{Q}_{t-1} \, \underline{m}_{t-1} \, .$$

From the definition of one step ahead forecast error e_t we know that

$$z_t = \underline{u}_t \, \underline{m}_{t-1} + e_t$$

so

$$\underline{Q}_t \, \underline{m}_t = \underline{u}_t' \, (\underline{u}_t \, \underline{m}_{t-1} + e_t) + \beta \underline{Q}_{t-1} \, \underline{m}_{t-1}$$

$$= (\underline{u}_t' \, \underline{u}_t + \beta \, \underline{Q}_{t-1}) \, \underline{m}_{t-1} + \underline{u}_t' \, e_t$$

Now $\underline{u}_t' \, \underline{u}_t + \beta \, \underline{Q}_{t-1} = \underline{Q}_t$ from (5.2.3.1)
so

$$\underline{m}_t = \underline{m}_{t-1} + \underline{Q}_t^{-1} \, \underline{u}_t' \, e_t$$

$$= \underline{m}_{t-1} + \underline{A}_t \, e_t \qquad (5.3.1.10)$$

as $\underline{A}_t = \underline{Q}_t^{-1} \, \underline{u}_t'$.

This completes the proof.

5.3.2 Corollaries

5.3.2.1 Corollary 1

For the time series the results (5.3.1.8) and (5.3.1.11) of the Fundamental Theorem of G.E.W.R. for the estimate $\underline{m}_t$ of $\underline{\theta}$ at time t, give

$$\underline{m}_t = \underline{Q}_t^{-1} \, \underline{H}_t \qquad (5.3.2.1.1)$$

where $\underline{Q}_t$ and $\underline{H}_t$ are as defined in (5.2.3.2) and (5.2.4.2), i.e.

$$\underline{Q}_t = \underline{u}_t' \underline{u}_t + \beta(\underline{G}')^{-1} \underline{Q}_{t-1} \underline{G}^{-1} \qquad \text{and}$$

$$\underline{H}_t = \underline{u}_t' z_t + \beta(\underline{G}')^{-1} \underline{H}_{t-1} \quad ;$$

and $$\underline{m}_t = \underline{G}\,\underline{m}_{t-1} + \underline{A}_t e_t \tag{5.3.2.1.2}$$

where $$e_t = z_t - d_t$$

$$= \psi_t(\beta^{1/2} B) y_t - \underline{u}_t \underline{G}\,\underline{m}_{t-1}$$

and $\underline{A}_t$, the updating or gain vector is calculated through the following recurrence relations.

Following D.E.W.R. (3.3) the recurrence relations for calculating $\underline{A}_t$ are

$$\underline{K}_t = \underline{G}\,\underline{Q}_{t-1}^{-1} \underline{G}' / \beta$$

$$\hat{Y}_t = 1 + \underline{u}_t \underline{K}_t \underline{u}_t'$$

$$\underline{A}_t = \underline{K}_t \underline{u}_t' (\hat{Y}_t)^{-1}$$

$$\underline{Q}_t^{-1} = (\underline{I} - \underline{A}_t \underline{u}_t) \underline{K}_t \tag{5.3.2.1.3}$$

For these recurrence relations, no matrix inversions are involved, but if required for any intermediate estimate, can be found from the above expression of $\underline{Q}_t$, by using the matrix identity (3.2).

By using estimate $\underline{m}_t$, one step ahead forecast $\hat{y}_t$ for the original or observed time series y_t is obtained from the derived series z_t as

$$\hat{y}_t = d_t - \{ z_t - y_t \}$$

$$= \underline{u}_t \underline{G}\ \underline{m}_{t-1} - \{\psi_t(\beta^{1/2} B)y_t - y_t\}. \qquad (5.3.2.1.4)$$

5.3.2.2 <u>Corollary 2</u>

Restricting the discount factor β such that $0 < \beta < \min\limits_i |\lambda_i^2|$, guaratees the existence of the following limits as $t \to \infty$, and shows that in practice E.W.R. is not to be applied to mixtures of high and low frequencies. From section (5.2.2) as $t \to \infty$

$$\text{Lim } \underline{u}_t = \underline{u} = \sum_{i=0}^{\infty} \alpha_i \underline{f}_i \qquad \text{a 1xn vector}$$

$$\text{Lim } \underline{Q}_t = \underline{Q} = \underline{u}' \underline{u} + \beta(\underline{G}')^{-1} \underline{Q}\ \underline{G}^{-1} \qquad \text{an n square matrix}$$

$$\text{Lim } \underline{A}_t = \underline{A} = \underline{Q}^{-1} \underline{u}' \qquad \text{an nx1 vector}$$

$$\underline{H}_t \to \underline{u}' z_t + \beta(\underline{G}')^{-1} \underline{H}_{t-1} \qquad \text{an nx1 vector}$$

$$d_t \to \underline{u}\ \underline{G}\ \underline{m}_{t-1} \qquad \text{a scalar}$$

$$\underline{m}_t = \underline{Q}_t^{-1} \underline{H}_t \to \underline{Q}^{-1}\underline{H}_t \quad \text{and} \quad \underline{m}_t = \underline{G}\ \underline{m}_{t-1} + \underline{A}_t e_t \to \underline{G}\ \underline{m}_{t-1} + \underline{A}\ e_t$$

<u>Comment</u>

The importance of the above restriction on the discount factor β is highlighted in theorem (T4)(5.3.5), where the effects of violation of this restriction are given.

5.3.3 Theorem (T3)

Consider at time t, a local model (5.2.3.2.1), i.e.

$$y_{t+i} = \underline{f}_i \, \underline{\theta} + \varepsilon_{t+i}$$

where the Coloured Noise $\varepsilon_t \sim$ ARMA (p,q) has representation $\phi_p(B)\varepsilon_t = \eta_q(B)\delta_t$ as defined in (4.2.0.1). Given a discount factor β such that

$$0 < \beta < \min_i |\lambda_i^2| \quad \text{and writing}$$

$$z_t = \psi_t(\beta^{1/2} B)\, y_t \ ;$$

the limiting result is

$$\lim_{t\to\infty} \{ \prod_{i=1}^{n} (1 - \lambda_i B)\, z_{t+1} - \prod_{i=1}^{n} (1 - \beta B/\lambda_i)\, e_{t+1} \} = 0 .$$

Proof:

Let $\underline{P}_1(B)$ and $\underline{P}_2(B)$ be matrices and $P_3(B)$ and $P_4(B)$ scalars all of whose elements are polynomials of order n in the backward shift operator B.

From corollary 1 of theorem (T2) (5.3.2) we know that

$$\underline{m}_t = \underline{G}\, \underline{m}_{t-1} + \underline{A}_t\, e_t .$$

Using the backward shift operator B we get

$$(\underline{I} - \underline{G} B)\, \underline{m}_t = \underline{A}_t\, e_t .$$

Assuming that $(\underline{I} - \underline{G} B)$ is invertible we can write

$$\underline{m}_t = (\underline{I} - \underline{G} B)^{-1} \underline{A}_t\, e_t . \qquad (5.3.3.1)$$

Now $\quad (\underline{I} - \underline{G} B)^{-1} = \{ \mathrm{adj}(\underline{I} - \underline{G} B) \} / |\underline{I} - \underline{G} B|$

$$= \{ \mathrm{adj}(\underline{I} - \underline{G} B) \} / \prod_{i=1}^{n} (1 - \lambda_i B)$$

Letting $\mathrm{adj}(\underline{I} - \underline{G}B) = \underline{P}_1(B)$ we can write (5.3.3.1) as

$$\underline{m}_t = \{ \underline{P}_1(B) / \prod_{i=1}^{n} (1 - \lambda_i B) \} \underline{A}_t e_t \qquad (5.3.3.2)$$

where the elements of $\underline{P}_1(B)$ are polynomials in B of degree n-1 as the elements of $\underline{P}_1(B)$ are minors of the elements of n square matrix $(\underline{I} - \underline{G}B)$. In order to obtain another relationship between the observations and the errors defining

$$x_{t+1} = \underline{u}\,\underline{G}\,\underline{m}_t + e_{t+1} \qquad (5.3.3.3)$$

and substituting here the above expression for $\underline{m}_t$ we get

$$x_{t+1} = u\,\underline{G}\,\{ \underline{P}_1(B) / \prod_{i=1}^{n} (1 - \lambda_i B) \} \underline{A}_t e_t + e_{t+1}.$$

Further substituting here the value of $\underline{A}_t = \underline{Q}_t^{-1} \underline{u}_t$ from (5.3.1.1) and re-arranging the terms we get

$$\prod_{i=1}^{n} (1 - \lambda_i B) x_{t+1} = \{ \underline{u}\,\underline{G}\,\underline{P}_1(B)\, \underline{Q}_t^{-1} \underline{u}_t' B + \prod_{i=1}^{n} (1 - \lambda_i B) \} e_{t+1}.$$

$$(5.3.3.4)$$

Now from (5.2.4.2.2)

$$\underline{H}_t = \beta(\underline{G}')^{-1} \underline{H}_{t-1} + \underline{u}_t' z_t$$

or $$(\underline{I} - \beta(\underline{G}')^{-1} B)\, \underline{H}_t = \underline{u}_t' z_t.$$

Assuming that the bracket expression is invertible, we can write

$$\underline{H}_t = (\underline{I} - \beta(\underline{G}')^{-1} B)^{-1} \underline{u}_t' z_t.$$

Now

$$(I - \beta(\underline{G}')^{-1}B)^{-1} = \{adj(I - \beta(\underline{G}')^{-1}B)\} / |\underline{I} - \beta(\underline{G}')^{-1}B|$$

$$= \{adj(\underline{I} - \beta(\underline{G}')^{-1}B)\} / \prod_{i=1}^{n} (1 - \beta B/\lambda_i).$$

Letting $adj(\underline{I} - \beta(\underline{G}')B) = \underline{P}_2(B)$ we can write the expression for $\underline{H}_t$ as

$$\underline{H}_t = \{ \underline{P}_2(B) / \prod_{i=1}^{n} (1 - \beta B/ \lambda_i) \} \underline{u}_t' z_t . \qquad (5.3.3.5)$$

where the elements of $\underline{P}_2(B)$ are polynomials in B of degree n-1 due to the similar reason of the degree of the matrix $\underline{P}_1(B)$.

Defining $w_{t+1} = \underline{u}\ \underline{G}\ \underline{Q}^{-1}\ \underline{H}_t + e_{t+1}$. (5.3.3.6)

Then substituting the value of $\underline{H}_t$ given in (5.3.3.5) it follows that

$$w_{t+1} = \underline{u}\ \underline{G}\ \underline{Q}^{-1}\{ \underline{P}_2(B) / \prod_{i=1}^{n} (1 - \beta B/\lambda_i)\} \underline{u}_t' z_t + e_{t+1}.$$

Re-arranging the terms we get

$$\prod_{i=1}^{n} (1 - \beta B/ \lambda_i) e_{t+1} = \prod_{i=1}^{n} (1 - \beta B/ \lambda_i) w_{t+1}$$

$$- \underline{u}\ \underline{G}\ \underline{Q}^{-1}\ \underline{P}_2(B)\ \underline{u}_t' z_t. \qquad (5.3.3.7)$$

Now $z_{t+1} = \underline{u}_{t+1} \underline{G}\ \underline{m}_t + e_{t+1}$ from the definition (5.3.2.1) of one step ahead forecast error. Hence, since $\lim_{t\to\infty} \underline{u}_{t+1} = \underline{u}$

$$\lim_{t\to\infty} (z_{t+1} - \underline{u}\ \underline{G}\ \underline{m}_t - e_{t+1}) = 0$$

and hence $\lim_{t\to\infty}(z_{t+1} - w_{t+1}) = 0$ showing that as $t\to\infty$, $z_{t+1} \to w_{t+1}$ (remembering that in the limit $t\to\infty$, from (5.3.2.2) $\underline{m}_t = \underline{Q}^{-1}\underline{H}_t$). Similarly as $t\to\infty$, $z_{t+1} \to x_{t+1}$.

Using these limiting results and writing

$$1 + BP_3(B) = \underline{u}\ \underline{G}\ \underline{P}_1(B)\ \underline{Q}^{-1}\ \underline{u}'B + \prod_{i=1}^{n} (1-\lambda_i B)$$

where $\underline{Q}^{-1}\underline{u}'$ is a limiting value of $\underline{Q}_t^{-1}\underline{u}_t'$

and $$1 + BP_4(B) = \prod_{i=1}^{n} (1-\beta B/\lambda_i) - \underline{u}\ \underline{G}\ \underline{Q}^{-1}\ \underline{P}_2(B)\ \underline{u}'\ B$$

we can write in the limit (5.3.3.4) and (5.3.3.7) as

$$\lim_{t\to\infty}\{ \prod_{i=1}^{n} (1-\lambda_i B)z_{t+1} - (1+BP_3(B))e_{t+1} \} = 0 \quad (5.3.3.8)$$

and $$\lim_{t\to\infty}\{ \prod_{i=1}^{n} (1- \beta B/\lambda_i)e_{t+1} - (1+BP_4(B))z_{t+1}\} = 0 \quad (5.3.3.9)$$

respectively.

We have now established two limiting equations, each relating the observation series and the error series. These limiting equalities are true for all values of the discount factor β such that $0<\beta<\min|\lambda_i^2|$. Furthermore the coefficients of B^i for $i = 0,1,\ldots,n$ in the two equations must be proportional. However taking the coefficients of z_{t+1} and e_{t+1} it is seen that the proportionality constant is 1 so that the equation must be identical. Hence equating the coefficients of these equalities in B^i, we get the required result

$$\lim_{t\to\infty} \left(\prod_{i=1}^{n} (1-\lambda_i B)z_{t+1} - \prod_{i=1}^{n} (1-\beta B/\lambda_i)e_{t+1} \right) = 0 \quad (5.3.3.10)$$

Comment

Apart from extending the results of McKenzie(1976) and the E.W.R. results of Godolphin-Harrison(1976), the proof is simple straightforward and shorter than that used for special cases. The theorem is not dependent upon the estimation method being optimal in any way. Hence the result is general. It gives equivalences between the limiting forecast functions derived from G.E.W.R., ARIMA and Constant Dynamic Linear Models.

5.3.4 Corollaries

5.3.4.1 Corollary 1

For an ARMA(p,q) type coloured noise process we know that $\psi(B) = \phi(B)/\eta(B)$, so writing

$$z_{t+1} = \psi(\beta^{1/2}B)y_{t+1} = \{\phi(\beta^{1/2}B)/\eta(\beta^{1/2}B)\}\ y_{t+1}$$

we can write the limiting result (5.3.3.9) as

$$\lim_{t\to\infty} \{\phi(\beta^{1/2}B) \prod_{i=1}^{n} (1-\lambda_i B)y_{t+1} - \eta(\beta^{1/2}B) \prod_{i=1}^{n} (1-\beta B/\lambda_i)e_{t+1} = 0 \qquad (5.3.4.1.1)$$

5.3.4.2 Corolloray 2

When $\psi(B) = 1$, then $z_{t+1} = y_{t+1}$ and the limiting result (5.3.3.9) reduces to

$$\lim_{t\to\infty} \{ \prod_{i=1}^{n} (1-\lambda_i B)y_{t+1} - \prod_{i=1}^{n} (1-\beta B/\lambda_i)e_{t+1} \} = 0 \qquad (5.3.4.2.1)$$

the E.W.R. case.

This special case of the theorem (T3) is also given by McKenzie (1976).

5.3.4.3 Corollary 3

If $\psi(B) = 1$ and $\underline{G} = \underline{J}_n(\lambda)$, the Jordan form for n equal eigenvalues λ, then the limiting result (5.3.3.9) reduces to

$$\lim_{t\to\infty} \{(1 - \lambda B)^n y_{t+1} - (1 - \beta B/\lambda)^n e_{t+1}\} = 0 \quad (5.3.4.3.1)$$

If $\lambda = 1$, then this expression further reduces to

$$\lim_{t\to\infty} \{ (1 - B)^n y_{t+1} - (1 - \beta B)^n e_{t+1} \} = 0 \quad (5.3.4.3.2)$$

a result given by Godolphin and Harrison (1975).

5.3.5 Theorem (T4)

For a local model (5.2.3.2.1), i.e.

$$Y_{t+i} = \underline{f}_i \ \underline{\theta} + \varepsilon_{t+i}$$

where Y_{t+i}, $\underline{f}_i$ and $\underline{\theta}$ have their usual interpretation, $\varepsilon_t \sim \text{ARMA}(p,q)$ and the discount factor β is such that

$$0 < \beta < \min|\lambda_i^2| \quad \text{for } i=1,\dots,m$$

and

$$\beta > \lambda_i^2 \quad \text{for } i=m+1,\dots,n$$

where $0 < \beta < 1$ and $0 < |\lambda_i| < 1$;

in the limit $t\to\infty$, ,we get a limiting result of a reduced order i.e.

$$\lim_{t\to\infty} \{ \prod_{i=1}^{m} (1-\lambda_i B) z_{t+1} - \prod_{i=1}^{m} (1-\beta B/\lambda_i) e_{t+1} \} = 0 .$$

Proof:

Writing $\underline{Q}_t^{-1} = \underline{C}_t$, $\underline{K}_t = \underline{R}_t$, the dynamic recurrence relations defined in the section (5.3.2) can be written for the model (5.3.5.1) as

$$\underline{R}_t = \underline{G}\ \underline{C}_{t-1}\ \underline{G}' / \beta$$

$$\hat{Y}_t = 1 + \underline{u}_t\ \underline{R}_t\ \underline{u}_t'$$

$$\underline{A}_t = \underline{R}_t\ \underline{u}_t' (\hat{Y}_t)^{-1}$$

$$\underline{C}_t = (\underline{I} - \underline{A}_t\ \underline{u}_t) \underline{R}_t$$

$$\hat{y}_t = d_t - (z_t - y_t)$$

$$e_t = z_t - d_t$$

$$\underline{m}_t = \underline{G}\ \underline{m}_{t-1} + \underline{A}_t\ e_t \qquad (5.3.5.2)$$

Let $\underline{f}_t = \underline{f} = (1,1,....,1)$ a 1xn vector

$$\underline{G} = \begin{bmatrix} \underline{\Lambda}_1 & \underline{0} \\ \underline{0} & \underline{\Lambda}_2 \end{bmatrix} \quad \text{an nxn matrix}$$

where $\underline{\Lambda}_1 = \mathrm{diag}(\lambda_1 , \lambda_2 , ... \lambda_m)$

$\underline{\Lambda}_2 = \mathrm{diag}(\lambda_{m+1}, \lambda_{m+2}, ... \lambda_n)$

$$\underline{S}_{t-1} = \begin{bmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{12} & \underline{S}_{22} \end{bmatrix}$$

where $\underline{S}_{11}$, $\underline{S}_{12}$ and $\underline{S}_{22}$ are sub-matrices of the appropriate dimensions matching with the corresponding dimensions of the sub-matrices of the $\underline{G}$ matrix

then, after t-step transitions or iterations, we can write the recursions for $\underline{R}_t$, $\hat{Y}_t$, $\underline{A}_t$, $\underline{S}_t$ and $\underline{C}_t$ as follows.

$$\underline{R}_t = \beta^{-t}\, \underline{G}^t\, \underline{S}_{t-1}\, \underline{G}^t$$

$$= \beta^{-t} \begin{bmatrix} \underline{\Lambda}_1 & \underline{0} \\ \underline{0} & \underline{\Lambda}_2 \end{bmatrix}^t \begin{bmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{12} & \underline{S}_{22} \end{bmatrix}_{t-1} \begin{bmatrix} \underline{\Lambda}_1 & \underline{0} \\ \underline{0} & \underline{\Lambda}_2 \end{bmatrix}^t$$

$$= \beta^{-t} \begin{bmatrix} \underline{\Lambda}_1^t \underline{S}_{11} \underline{\Lambda}_1^t & \underline{\Lambda}_1^t \underline{S}_{12} \underline{\Lambda}_2^t \\ \underline{\Lambda}_2^t \underline{S}_{12} \underline{\Lambda}_1^t & \underline{\Lambda}_2^t \underline{S}_{22} \underline{\Lambda}_2^t \end{bmatrix}_{t-1} = \begin{bmatrix} \underline{R}_{11} & \underline{R}_{12} \\ \underline{R}_{12} & \underline{R}_{22} \end{bmatrix}_t \tag{5.3.5.4}$$

$$\hat{Y}_t = 1 + \underline{u}_t\, \underline{R}_t\, \underline{u}_t' \tag{5.3.5.5}$$

$$\underline{A}_t = \underline{R}_t\, \underline{u}_t'\, (\, \hat{Y}_t\,)^{-1} = \begin{pmatrix} \underline{A}_1 \\ \underline{A}_2 \end{pmatrix}_t \tag{5.3.5.6}$$

where $\underline{A}_1 = (\, A_1 \;, \ldots, \; A_m \,)_t'$

$\underline{A}_2 = (\, A_{m+1}, \ldots, \; A_n \,)_t'$

$$\underline{S}_t = (\, \underline{I} - \underline{G}^{-t} \underline{A}_t\, \underline{u}_t\, \underline{G}^t \,)\, \underline{S}_{t-1} \tag{5.3.5.7}$$

or $$\underline{S}_t = \begin{bmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{12} & \underline{S}_{22} \end{bmatrix}_t$$

$$\underline{C}_t = \beta^{-t}\, \underline{G}^t\, \underline{S}_t\, \underline{G}^t$$

$$= \beta^{-t} \begin{bmatrix} \underline{\Lambda}_1^t \underline{S}_{11} \underline{\Lambda}_1^t & \underline{\Lambda}_1^t \underline{S}_{12} \underline{\Lambda}_2^t \\ \underline{\Lambda}_2^t \underline{S}_{12} \underline{\Lambda}_1^t & \underline{\Lambda}_2^t \underline{S}_{22} \underline{\Lambda}_2^t \end{bmatrix}_t = \begin{bmatrix} \underline{C}_{11} & \underline{C}_{12} \\ \underline{C}_{12} & \underline{C}_{22} \end{bmatrix}_t . \tag{5.3.5.8}$$

Now by definition, the discount factor

$$\beta < |\lambda_i^2| \qquad \text{for } i=1,\ldots, m$$

$$\beta > |\lambda_i^2| \qquad i=m+1,\ldots, n$$

so in the limit $t \to \infty$

$$\underline{C}_{22,t} = \beta^{-t}\, \underline{\Lambda}_2^t\, \underline{S}_{22}\, \underline{\Lambda}_2^t \to \underline{0} \tag{5.3.5.9}$$

$$\underline{C}_{12,t} = \beta^{-t}\, \underline{\Lambda}_2^t\, \underline{S}_{12}\, \underline{\Lambda}_1^t \to \underline{0} \tag{5.3.5.10}$$

$$\underline{C}_{11,t} = \beta^{-t}\, \underline{\Lambda}_1^t\, \underline{S}_{11}\, \underline{\Lambda}_1^t \to \hat{\underline{C}}_{11} \tag{5.3.5.11}$$

The first two limits (5.3.5.9) and (5.3.5.10) are quite obvious. The third (5.3.5.11) is achieved by assuming that

i) $\underline{\theta}_t, \underline{m}_t \in \Omega$ (5.3.5.12)

where Ω is some compact parameteric space on σ-field.

ii) $E\,|\,Y_t - \overset{0}{Y}_{s,t}\,|^4 < c\ \beta^{t-s}$ (5.3.5.13) and $$E \left| \underline{\theta}_t - \overset{0}{\underline{\theta}}_{s,t} \right|^4 < c\,\beta^{t-s} \qquad (5.3.5.14)$$

for some constant $c < \infty$ and $0 < \beta < 1$, $s \leq t$

where $\overset{0}{Y}_{s,t}$, $\overset{0}{\underline{\theta}}_{s,t}$ are some random variables that belong to Ξ^t (σ- algebra generated by White Noise δ_t,etc.) but are independent of Ξ^s.

iii) The model is completely observable and controllable

i.e. $\underline{T} = (\underline{f} , \underline{f}\,\underline{G} , \ldots , \underline{f}\,\underline{G}^{n-1})$ is of full rank. (5.3.5.16)

For any dynamic model that fulfils the stated conditions the limit

$$\sup_{\underline{\theta}_t \in \Omega} \left| \underline{C}_t - \hat{\underline{C}}_t \right| \to \underline{0} \quad \text{as } t \to \infty \qquad (5.3.5.17)$$

exists with probability 1 as shown by Ljung (1978),

where $\hat{\underline{C}}_t = E (\underline{C}_t)$

This result is valid for all the dynamic systems which are exponentially stable, i.e. where the remote past decays exponentially.

Following this result we see that

$$\sup_{\underline{\theta}_t \in \Omega} \left| \underline{C}_{11,t} \right| \to \hat{\underline{C}}_{11} \qquad (5.3.5.18)$$

Hence in the limit $t \to \infty$

$$\underline{C}_t \to \hat{\underline{C}} = \begin{bmatrix} \hat{\underline{C}}_{11} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} \qquad (5.3.5.19)$$

$$\hat{Y}_t \rightarrow \hat{Y} \tag{5.3.5.20}$$

$$\underline{A}_t \rightarrow \underline{A} = \begin{pmatrix} \underline{A}_1 \\ \underline{0} \end{pmatrix} \tag{5.3.5.21}$$

Now $\underline{m}_t = \underline{G}\ \underline{m}_{t-1} + \underline{A}_t\ e_t$

or
$$\begin{pmatrix} \underline{m}_1 \\ \underline{m}_2 \end{pmatrix}_t = \begin{bmatrix} \underline{\Lambda}_1 & \underline{0} \\ \underline{0} & \underline{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \underline{m}_1 \\ \underline{m}_2 \end{bmatrix}_{t-1} + \begin{bmatrix} \underline{A}_1 \\ \underline{A}_2 \end{bmatrix}_t e_t \tag{5.3.5.22}$$

From the limiting results stated above, it follows that

$$\underline{m}_{1,t} \rightarrow \underline{\Lambda}_1\underline{m}_{1,t-1} + \underline{A}_1\ e_t$$

$$\underline{m}_{2,t} \rightarrow \underline{0}$$

thus reducing the original system of order n to a system of order m.

Hence from theorem (T3) (5.3.3), in the limit $t \rightarrow \infty$ we get

$$\left\{ \prod_{i=1}^{m} (1 - \lambda_i B) z_{t+1} - \prod_{i=1}^{m} (1 - \beta B/\lambda_i)\ e_{t+1} \right\} = 0 \tag{5.3.5.23}$$

Comments

This theorem highlights the importance of the restriction on the discount factor, i.e. $0 < \beta < \min|\lambda_i^2|$ if the same degree of polynomial model is required throughout the analysis of the time series. If the stated restriction on the discount factor β is violated or we do not fully comply with it, i.e. the discount factor is selected such that

$$\beta < |\lambda_i^2| \qquad \text{for } i=1,\ldots\ldots, m$$

and $\beta > |\lambda_i^2|$ for $i=m+1,\ldots\ldots, n$

then as is clear from the above theorem (T4) that in the limit as $t \to \infty$ the polynomial of degree n fitted to the time series reduces to degree m. Such a setting of β may be useful in a restricted class of models, where for some reason a higher degree polynomial is required for a short span and a polynomial of reduced degree is required when t becomes large.

For example if a second order model

$$Y_t = \underline{f}\ \underline{\theta}_t + v_t$$

$$\underline{\theta}_t = \underline{G}\ \underline{\theta}_{t-1} + \underline{w}_t$$

where $v_t \sim N(\ 0,\ V\ )$, $\underline{w} \sim N(\ \underline{0},\ \underline{W}\ )$

and $\underline{f} = (1\ \ 1)$, $\underline{G} = \mathrm{diag}(\lambda_1\ ,\ \lambda_2)$

is fitted to represent a trend or low frequency and β is selected such that

$$|\lambda_2^2| < \beta < |\lambda_1^2|$$

then in the begining the model will behave like a second order or degree model which, generally is sufficient to capture most of the movements in the low frequency and as $t \to \infty$, the order of the model will be reduced to one. In the limit $t \to \infty$

$$\underline{A}_t = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}_t \to \begin{pmatrix} A_1 \\ 0 \end{pmatrix}$$

and the estimator

$$\underline{m}_t = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}_t \to \begin{pmatrix} \lambda_1 m_{1,t-1} + A_1\ e_t \\ 0 \end{pmatrix}$$

Otherwise in the case $0 < \beta < \min_i |\lambda_i^2|$

$$\underline{m}_t = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}_t \to \begin{pmatrix} \lambda_1 m_{1,t-1} + A_1\ e_t \\ \lambda_2 m_{2,t-1} + A_2\ e_t \end{pmatrix}$$

5.3.6 Characterisation Theorem (T5)

Defining:

$E_{t-i} = y_{t-i} - \underline{f}_1 \underline{m}_{t-1-i}$ as the 'trend' or low frequency errors for i=0,1... ;

$$\alpha(B) = \sum_{i=0}^{\infty} \alpha_i B^i \quad ,$$

$$v(B) = \sum_{i=0}^{\infty} v_i B^i$$

where
$$v_k = \begin{cases} 1 & \text{if } k = 0 \\ \underline{f}_0 \sum_{i=k}^{\infty} \alpha_i \underline{G}^{-(i-k)} \underline{A} & \text{if } k > 0 \end{cases}$$

and $E_k = e_k = 0 \qquad$ for $k \leq 0$

then for the sequence E_t

$$\alpha(B) E_t = v(B) e_t$$

Proof:

Let the residuals e^*_{t-i} be defined such that

$$y_{t-i} = \underline{f}_{1-i} \underline{m}_{t-1} + e^*_{t-i}$$

then
$$e^*_{t-i} = y_{t-i} - \underline{f}_{1-i} \underline{m}_{t-1}. \qquad (5.3.6.1)$$

Adding and taking away $\underline{f}_1 \underline{m}_{t-1-i}$ we get

$$e^*_{t-i} = (y_{t-i} - \underline{f}_1 \underline{m}_{t-1-i}) - (\underline{f}_{1-i} \underline{m}_{t-1} - \underline{f}_1 \underline{m}_{t-1-i}).$$

or $\quad e^*_{t-i} = E_{t-i} - \underline{f}_1 \, (\, \underline{G}^{-i} \, \underline{m}_{t-1} - \underline{m}_{t-1-i} \,)$ as $\underline{f}_{-i} = \underline{f}_0 \underline{G}^{-i}$.

Now in the limit $t \to \infty$ (5.3.6.2)

$$\underline{m}_{t-1} = \underline{G} \, \underline{m}_{t-2} + \underline{A} \, e_{t-1}$$

which in a recurrence form can be written as

$$\underline{m}_{t-1} = \sum_{j=0}^{i-1} \underline{G}^j \, \underline{A} \, e_{t-1-j} + \underline{G}^i \, \underline{m}_{t-1-i} \, . \qquad (5.3.6.3)$$

Substituting in (5.3.6.2), we get

$$e^*_{t-i} = E_{t-i} - \underline{f}_1 \Big[\, \underline{G}^{-i} \Big\{ \sum_{j=0}^{i-1} \underline{G}^j \, \underline{A} \, e_{t-1-j} + \underline{G}^i \underline{m}_{t-1-i} \Big\} - \underline{m}_{t-1-i} \Big]$$

$$= E_{t-i} - \underline{f}_1 \sum_{j=0}^{i-1} \underline{G}^{j-i} \, \underline{A} \, e_{t-1-j} \, . \qquad (5.3.6.4)$$

Now

$$\begin{aligned} e_t &= z_t - d_t \\ &= \alpha_t(B) \, y_t - \sum_{i=0}^{t-1} \alpha_i \, \underline{f}_{-i} \, \underline{G} \, \underline{m}_{t-1} \\ &= \sum_{i=0}^{t-1} \alpha_i \, (\, y_{t-i} - \underline{f}_{1-i} \, \underline{m}_{t-1} \,) \\ &= E_t + \sum_{i=1}^{t-1} \alpha_i \, e^*_{t-i} \, . \end{aligned}$$

Substituting here the expression (5.3.6.4) for e^*_{t-i} and re-arranging the terms, we get

$$e_t = \sum_{i=0}^{t-1} \alpha_i E_{t-i} - \underline{f}_1 \sum_{i=1}^{t-1} \sum_{j=0}^{i-1} \alpha_i \underline{G}^{j-i} \underline{A}\, e_{t-1-j} \tag{5.3.6.5}$$

Now as $t \to \infty$ we can write

$$e_t = \alpha(B)E_t - \underline{f}_1 \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \alpha_i \underline{G}^{j-i} \underline{A}\, e_{t-1-j}$$

$$= \alpha(B)E_t - \underline{f}_1 \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} \alpha_i \underline{G}^{k-i-1} \underline{A}\, e_{t-k}$$

or

$$\alpha(B)E_t = e_t + \underline{f}_1 \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} \alpha_i \underline{G}^{k-i-1} \underline{A}\, e_{t-k}$$

$$= v(B)\, e_t . \tag{5.3.6.6}$$

which completes the proof.

<u>Corollary</u>

If $\psi(B) = \sum_{i=0}^{p} \psi_i B^i$ then if the $e_t \sim (0, \sigma^2)$ are independent for $t > 0$, the sequence $E_t \sim$ ARMA(p,p) as $t \to \infty$

<u>Comment</u>:

In characterisation theorem (T5), no assumption about the nature (other than stationarity) of the e_t series is made. In general the E_t series has an ARMA interpretation if initially (prior to filtration) the e_t series is an ARMA process or after filtration, by some appropriate forecast model, e_t forms a white noise sequence, such that in the limit $t \to \infty$

$$\text{Lim } E(e_t) = 0 = \text{Lim } E(e_t\, e_{t+k})$$

and $\quad Var(e_t) = \sigma^2$

For example, consider $(1-B)^n y_t = (1-\beta B)^n e_t$. Then if initially $e_t \sim$ ARMA(p,q), the series $y_t \sim$ ARIMA(p, n, n+q) and after filtration $y_t \sim$ ARIMA(0,n,n) if $e_t \sim (0,\sigma^2)$ are uncorrelated,i.e. series of one step ahead forecast errors generated by some appropriate forecast model are such that they form a white noise sequence with mean zero and variance σ^2.

5.4 Practical Implications

In this section the use of the theorems (presented in previous section) is illustrated, especially with reference to corollaries (5.3.4.2) and (5.3.4.3). For elaboration, some more theorems concerned with practical implications are given. First a theorem relevant only to distinct eigenvalues, concerning the updating adaptive vector $\underline{A}_t$ in the recurrence

$$\underline{m}_t = \underline{G}\,\underline{m}_{t-1} + \underline{A}_t\, e_t$$

is given. Historically this theorem is due to Dobbie (1963), who presented it without mentioning the important restriction on the discount factor β .

5.4.1 Theorem (T6)

If $\underline{G} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ where the λ_i are all distinct and non zero, $\underline{f}_t = \underline{f}_o = (1,1,\ldots\ldots,1)$ and $0 < \beta < \min|\lambda_i^2|$ then

$$\lim_{t\to\infty} \underline{A}_t = \underline{A} = (A_1, A_2, \ldots\ldots, A_n)'$$

where $\quad A_i = (1-\beta/\lambda_i^2) \prod_{\substack{j=1 \\ j\neq i}}^{n} (1-\beta/\lambda_i\lambda_j)/(1-\lambda_j/\lambda_i)$

(5.4.1.1)

Comment

The theorem is of some practical value when the λ_i are complex and lie close to the unit circle, or on it. It is of limited value otherwise. For example with two distinct real eigenvalues

$$A_1 = (\lambda_1^2 - \beta)(\lambda_1\lambda_2 - \beta)/\{\lambda_1^2\lambda_2\ (\lambda_1 - \lambda_2)\}$$

$$A_2 = (\lambda_2^2 - \beta)(\lambda_1\lambda_2 - \beta)/\{\lambda_1\lambda_2^2\ (\lambda_2 - \lambda_1)\}$$

so that the ratio

$$A_1/A_2 = -\ \lambda_2(\lambda_1^2 - \beta)/\{\lambda_1(\lambda_2^2 - \beta)\} \quad .$$

Hence if $\lambda_1 > \lambda_2 > 0$ and $A_1 > 0 > A_2$

then as $\lambda_1 \to \lambda_2$, $A_1 \to \infty$ and $A_2 \to -\infty$, which together with the restriction that $0 < \beta < \min\{|\lambda_1^2|\ ,\ |\lambda_2^2|\}$ recalls the limited practical value of E.W.R.to cases of distinct real eigenvalues. The highest frequency restricts the choice of dicount factor. For example, let $\underline{G} = \text{diag}(\lambda_1\ ,\ \lambda_2)$ with $\lambda_1 = 1$ and $\lambda_2 = \beta^{1/2}\phi = 0.5$ and $\underline{f} = (1 \quad 1)$, then according to the restriction on β, the value of β should be less than 0.25, which is against the spirit of E.W.R. as for trend we need high value of β . With low value of β, the trend won't be clear.

Further analogous to case of experimental design, there is a degree of confounding of their effects, i.e. it is impossible to separate their effects, as revealed through the observed time series, but ofcourse this remark applies to distinct eigenvalues in general time series models.

In certain cases where economic or business activity can adequately be represented by exponential growth curves where the growth in the limit $t \to \infty$ reaches a saturation level (see section (7.5) of chapter seven) and λ_2 is selected to represent the growth such that $|\lambda_2^2| < \beta$ then as $t \to \infty$ $A_2 \to 0$, thus reducing the order of the original model from degree 2 to degree 1 as clear from theorem (T4)(5.3.5). This highlights the consequences of not fully complying with the stated restriction on β . For related discussion see Harrison-Akram (1983) and Akram-Harrison (1983).

5.4.2 Theorem (T7)

Let the transition matrix be $\underline{G} = \underline{J}_n(\lambda)$ a Jordan block for n equal eigenvalues, where $|\lambda| \leq 1$, then for an E.W.R. type D.L.M. (3.5.2.1), i.e.

$$Y_t = \underline{f}\ \underline{\theta}_t + v_t \quad ; \quad v_t \sim N(0, V)$$

$$\underline{\theta}_t = \underline{J}_n(\lambda)\ \underline{\theta}_{t-1} + \underline{w}_t \quad ; \quad \underline{w}_t \sim N(\underline{0}, \underline{W})$$

for which the k-steps ahead forecast function and $\underline{W} = (W_1, \ldots, W_n)$ are defined by (3.5.2.2) and (3.5.2.3) respectively,

i) the forecast function converges (in probability) to that derived from E.W.R. using a discount factor β .

ii) in the limit $t \to \infty$

$$\mathrm{Var}(e_t) = \hat{Y}_t \to \hat{Y} = (\lambda^2/\beta)^n\, V$$

and

iii) the Bayesian updating $(\theta_t | D_t) \sim N(\underline{m}_t, \underline{C}_t)$ with

$$\underline{m}_t = \underline{J}_n(\lambda)\ \underline{m}_{t-1} + \underline{A}_t\, e_t$$

is such that

$$\lim_{t \to \infty} \underline{A}_t = \underline{A} = (A_1, \ldots, A_n)'$$

where $A_1 = 1 - (\beta/\lambda^2)^n$,

$$A_{k+1} = \binom{n}{k} \alpha^k - \lambda\, A_k$$

and

$$A_n = \alpha^n / \lambda$$

where $\alpha = (\lambda^2 - \beta)/ \lambda$.

<u>Proof:</u>

(i) and (ii) are special cases of theorem (T3) (5.3.3) given in its corollaries. (ii) is a result (3.4.2.2.15) when $\lambda_i = \lambda$ for all i .

iii) For the Jordan block

$$\underline{J}_n(\lambda) = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \underline{0} \\ \underline{0} & & & \ddots & \\ & & & & 1 \\ & & & & \lambda \end{bmatrix}$$

we have $\underline{f} = (1,0,\ldots,0)$, which gives

$$\underline{f}\, \underline{J}_n(\lambda)\, \underline{m}_t = \lambda\, m_{1,t} + m_{2,t} .$$

Now the one step ahead observation at time t is

$$y_{t+1} = \underline{f}\, \underline{J}_n(\lambda)\, \underline{m}_t + e_{t+1}$$

$$= \lambda\, m_{1,t} + m_{2,t} + e_{t+1} \qquad (5.4.2.1)$$

also $\underline{m}_t = \underline{J}_n(\lambda)\, \underline{m}_{t-1} + \underline{A}_t\, e_t$

$$= (\underline{I} - \underline{J}_n(\lambda)\, B)^{-1}\, \underline{A}_t\, e_t \qquad (5.4.2.2)$$

assuming that $(\underline{I} - \underline{J}_n(\lambda)\, B)$ is invertible. Substituting above we get

$$y_{t+1} = (\underline{f}\, \underline{J}_n(\lambda)\, \{ \underline{I} - \underline{J}_n(\lambda)\, B \}^{-1}\, \underline{A}_t\, B + 1)\, e_{t+1}$$

$$= (1 + \underline{f}\, \underline{J}_n(\lambda)\, \{ \underline{P}_1(B)/(1-\lambda B)^n \}\, \underline{A}_t B)\, e_{t+1}$$

(5.4.2.3)

where $\underline{P}_1(B)$ is a (nxn) matrix whose elements are polynomial of order n in the backward shift operator B, such that

$$\underline{P}_1(B) = (\underline{I} - \underline{J}_n(\lambda)B)^{-1} (1 - \lambda B)^n. \qquad (5.4.2.4)$$

Defining

$$\underline{P}_1(B) = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ & & & \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \qquad (5.4.2.5)$$

where

$$p_{i,j} = \begin{cases} (1 - \lambda B)^{n-1} & \text{for } i=j;\ i,j=1,\dots,n \\ B^{j-i}(1 - \lambda B)^{n+i-j-1} & \text{for } j > i \\ 0 & \text{for all } j < i \end{cases}$$

we can write the expression (5.4.2.3) as

$$(1-\lambda B)^n y_{t+1} = \Big((1-\lambda B)^n + \sum_{j=1}^{n} \{ B^{j-1}(1-\lambda B)^{n-j} + B^{j-2}(1-\lambda B)^{n+1-j} \} B A_j \Big) e_{t+1} .$$

Writing $B^{-1}A_j = A_{j+1}$, considering $A_{n+1} = 0$ and re-arranging the terms, we get

$$(1-\lambda B)^n y_{t+1} = \Big(\sum_{j=0}^{n} (\lambda A_j + A_{j+1})(1-\lambda B)^{n-j} B^j \Big) e_{t+1} \qquad (5.4.2.6)$$

$$= (1 - \beta B/\lambda)^n e_{t+1} \quad \text{if and only if}$$

$\lambda A_k + A_{k+1} = \binom{n}{k} \alpha^k$. This completes the proof.

6

G.E.W.R AND DYNAMIC MODEL REPRESENTATION

6.1 Introduction

In this chapter I shall mostly be concerned with the construction of dynamic models for ARMA type coloured noise processes. An Autoregressive form of G.E.W.R. is presented as in practice quite often ARMA processes with invertible Moving Average part can be adequately modelled as parsimonious finite AR processes. For related discussion see Andel (1981). The way in which E.W.R. models can be extended to G.E.W.R. models is shown and illustrated with respect to polynomial trends. However, the models which are built this way are rarely in a desirable form for operation. Hence an important topic of reparameterisation and similar models is discussed and a method of transforming dynamic system to a desired parametric form is given. Recurrence relations for the transformation matrices that transform a dynamic system of a canonical form to a diagonal form and vice-versa are given.

In section seven, a state space representation of G.E.W.R. is presented, alongwith the construction procedure. At the end an On-Line Bayesian Learning Procedure for the variance V_t of the Coloured Noise processes is given.

6.2 Autoregressive G.E.W.R.

An ARMA process with Moving Average Operator

$$\eta(B) = \prod_{i=1}^{q} (1- \eta_i B)$$

can be inverted to give an infinite Autoregressive representation. In practice for high frequencies generated by an ARMA type Coloured Noise process the η_i lie well within the unit circle. Hence as is well known such ARMA processes can be adequately modelled as parsimonious finite AR processes; then we consider

$$\psi(B) = \phi(B) = \prod_{i=1}^{p} (1- \phi_i B) = \sum_{i=0}^{p} \psi_i B^i$$

with $\psi_0 = 1$ and $\psi_p \neq 0$.

In this case for $\alpha_i = \psi_i \beta^{i/2}$ we define

$$\alpha(B) = \prod_{i=1}^{p} (1- a_i B) = \sum_{i=0}^{p} \alpha_i B^i = \psi(\beta^{1/2} B).$$

The corollary of theorem (T5) (5.3.6) states that

$$\alpha(B) E_t = v(B) e_t$$

so that if $e_t \sim (0, \sigma^2)$ are independent, the deviations of the observations from the one step ahead predicted trend values are represented as an ARMA(p,p) process.

6.2.1 Theorem (T8)

Defining:

$\underline{J} = \{ j_{k,\ell} \}$ as an (n+p) square matrix of full rank,

$$\text{where} \quad j_{k,\ell} = \begin{cases} a_i & \text{if } k=\ell=i \ ; i=1,\ldots\ldots, p \\ \lambda_i & \text{if } k=\ell=i+p;\ i=1,\ldots\ldots, n \\ 1 & \text{if } k=i \ , \ell=i+1 \ ; i=1,\ldots,(n+p-1) \\ 0 & \text{otherwise} \end{cases}$$

and $\underline{f} = (1,0,\ldots\ldots,0)$ a (n+p) vector.

The Constant Dynamic Linear Model (defined over quadruple($\underline{f}$, $\underline{J}$, V, $\underline{W}$))

$$Y_t = \underline{f}\ \underline{\theta}_t$$

$$\underline{\theta}_t = \underline{J}\ \underline{\theta}_{t-1} + \underline{w}_t \qquad (6.2.1.1)$$

where $\underline{w}_t \sim N(\underline{0}, \underline{W})$ and $\underline{W} = \{W_{ij}\}$ is of rank n

with $W_{ij} = W_{ji} = 0$ for i=1,.....,(p-1)

is such that

$$\alpha(B) \prod_{i=1}^{n} (1-\lambda_i B)\, y_t = \prod_{i=1}^{n} (1-\gamma_i B)\, \rho_t$$

where $\rho_t \sim N(0,\sigma^2)$ are independent random variables and $\lim_{t\to\infty} e_t = \rho_t$.

The proof of this theorem is standard for constant Dynamic Linear Models (see Harrison (1967)).

6.2.2 Comments

The E.W.R. results of section (5.4.2) for the variance settings for constant D.L.M.'s now carry across to G.E.W.R. for the AR(p) case simply by using the variance settings with

$$W_{pj} = \begin{cases} V & \text{if } j=p \\ W_{jp} = 0 & \text{if } j\neq p \end{cases} \qquad (6.2.2.1)$$

and putting the remaining block as the W parameter variance for the E.W.R. case corresponding to $\underline{G} = \{g_{ij}\}$ where

$$g_{ij} = \begin{cases} \lambda_i & \text{if } j=i \\ 1 & \text{if } j=i+1 \\ 0 & \text{otherwise} \end{cases} \qquad (6.2.2.2)$$

However, although this gives an easy way of building an appropriate G.E.W.R. Dynamic Linear Model based on E.W.R. principles, it is probably not always suitable for operation and reparameterisation is required. This can easily be accomplished using the results of section (6.6).

It may be noted that for all these AR(p) G.E.W.R. models

$$\sigma^2 = V \prod_{i=1}^{n} (\lambda_i^2 / \beta) = \hat{Y} \qquad (6.2.2.3)$$

ii) The k-steps ahead forecast function for the model (6.2.1.1) is

$$F_t(k) = \underline{f}\ \underline{J}^k\ \underline{m}_t \qquad \text{for } k=1,2,\ldots..$$

and the recurrence relations for updating $\underline{m}_t$, an estimator of $\underline{\theta}_t$ are

$$\underline{R}_t = \underline{J}\ \underline{C}_{t-1}\ \underline{J}' + \underline{W}$$

$$\underline{A}_t = \underline{R}_t\ \underline{f}' /(\underline{f}\ \underline{R}_t\ \underline{f}')$$

$$\underline{C}_t = (\underline{I} - \underline{A}_t\ \underline{f})\ \underline{R}_t$$

$$\underline{m}_t = \underline{J}\ \underline{m}_{t-1} + \underline{A}_t\ e_t$$

$$e_t = y_t - \underline{f}\ \underline{J}\ \underline{m}_{t-1} \qquad (6.2.2.4)$$

If V associated with the $\underline{W}$ matrix is unknown at time t, then it is estimated On-Line through the Bayesian learning process given in section (3.4.4), replacing $\underline{G}$ by $\underline{J}$ and setting $\underline{f}$ accordingly.

6.2.3 Example

For n=2 and p=2 we have an AR(2) type G.E.W.R. Constant D.L.M.

$$Y_t = \underline{f}\ \underline{\theta}_t$$

$$\underline{\theta}_t = \underline{J}\ \underline{\theta}_{t-1} + \underline{w}_t \quad ; \quad \underline{w}_t \sim N(\underline{0}, \underline{W})$$

where $\underline{f} = (1,0,0,0)$

$$\underline{\theta}_t = (\theta_1, \theta_2, \theta_3, \theta_4)'_t$$

$$\underline{J} = \begin{bmatrix} \phi_1\beta^{1/2} & 1 & 0 & 0 \\ 0 & \phi_2\beta^{1/2} & 1 & 0 \\ 0 & 0 & \lambda_1 & 1 \\ 0 & 0 & 0 & \lambda_2 \end{bmatrix} \qquad (6.2.3.1)$$

$$\underline{W} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & V & 0 & 0 \\ 0 & 0 & W_1 & 0 \\ 0 & 0 & 0 & W_2 \end{bmatrix} \qquad (6.2.3.2)$$

where

$$W_1 = (1-\beta)(\lambda_1 + \lambda_2)(\lambda_1\lambda_2 - \beta)\ V/(\lambda_2\beta)$$

$$W_2 = (1-\beta)(\lambda_1\lambda_2 - \beta)(\lambda_1 - \lambda_2\beta)(\lambda_2^2 - \beta)\ V/(\lambda_2\beta^2) \qquad (6.2.3.3)$$

if V is known.

In the case where V is not known at time t, we consider $\underline{W}_t$ in place of $\underline{W}$ where

$$\underline{W}_t = V_t \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \dot{W}_1 & 0 \\ 0 & 0 & 0 & \hat{W}_2 \end{bmatrix} \qquad (6.2.3.4)$$

and $\dot{W}_1$ and $\dot{W}_2$ are the W_1 and W_2 of (6.2.3.3) without the variance component V.

If a diagonal form of the dynamic system is considered, then

$$\underline{f} = (1,1,1,1) \quad \text{and} \quad \underline{J} = \text{diag}(\ \phi_1\beta^{1/2}\ ,\ \phi_2\beta^{1/2}\ ,\ \lambda_1, \lambda_2\ )$$

In such a case the matrix $\underline{W}$ will be trasformed to $\underline{W}_1$ as

$$\underline{W}_1 = \underline{H}\ \underline{W}\ \underline{H}_1'$$

where $\underline{H}$ is the transformation matrix introduced in section (6.6).

6.3 The Polynomial AR(p) and Damped Polynomial AR(p) Dynamic Linear Model

For the model (6.2.1.1), we consider the case in which $\lambda_i = \lambda$ for $i=1,\ldots, n$ and $|\lambda| < 1$.

6.3.1 Theorem (T9)

The constant D.L.M. of AR(p) G.E.W.R. (6.2.1.1) in which $\underline{W} = \{ W_{ij} \}$ is such that

$$W_{ij} = \begin{cases} W_{i-p} & \text{for } j=i\ ;\ i=p,\ldots,(n+p) \\ 0 & \text{otherwise} \end{cases}$$

with $\quad W_k = \binom{n}{k} c^k\, W_0 \quad$ for $k=1,2,\ldots$

where

$$c = (\ \lambda^2 - \beta\ )(\ 1-\beta\ )/\ \beta \quad > 0$$

has

i) a limiting forecast function equivalent to that derived by applying G.E.W.R. with discount factor β and a precision matrix $\underline{P}_t$ for an AR(p) representation with

$$\psi(B) = \sum_{i=0}^{p} \alpha_i \left(\beta^{1/2} B \right)^i = \sum_{i=0}^{p} \psi_i B^i \qquad (6.3.1.1)$$

ii) a limiting one step ahead variance σ^2 such that

$$\sigma^2 = \mathrm{var}(e_t) = \hat{Y} = (\lambda^2/\beta)^n W_0 \qquad (6.3.1.2)$$

<u>Proof:</u>

Defining

$$x_t = \alpha(B)(1- \lambda B)^n y_{t+p+n} \qquad (6.3.1.3)$$

the Autocovariance Generating Function (A.C.G.F.) $\gamma_x(B)$ of x is such that

a) from the D.L.M.

$$\gamma_x(B) = \sum_{k=0}^{n} \left\{ (1- \lambda B)(1- \lambda B^{-1}) \right\}^{n-k} W_k$$

$$= \sum_{k=0}^{n} \left\{ - \lambda B^{-1} + (1+\lambda^2) - \lambda B \right\}^{n-k} W_k$$

$$(6.3.1.4)$$

and

b) from the G.E.W.R. representation

$$\gamma_x(B) = \left\{(1- \beta B/\lambda)(1- \beta B^{-1}/\lambda)\right\}^n \mathrm{var}(e_t)$$

$$\left\{- \lambda B^{-1} + (\beta+ \lambda^2/\beta) - \lambda B \right\}^n (\beta/\lambda^2)^n \sigma^2$$

$$(6.3.1.5)$$

c) equating coefficients of powers of B in (6.3.1.4) and (6.3.1.5), we get

$$W_0 = (\beta/\lambda^2)^n \sigma^2 = (\beta/\lambda^2)^n \hat{Y}$$

or

$$\hat{Y} = (\lambda^2/\beta)^n W_0 . \qquad (6.3.1.6)$$

Also from section (3.4.2) we know that for $\lambda_i = \lambda$

$$\hat{Y} = (\lambda^2/\beta)^n V. \qquad (6.3.1.7)$$

Comparing with (6.3.1.6) we see that $W_0 = V$

d) Substituting $W_k = \binom{n}{k} c^k W_0$ in (a) we get

$$\gamma_x(B) = \sum_{k=0}^{n} \binom{n}{k} c^k \left[-\lambda B^{-1} + (1+\lambda^2) - \lambda B \right]^{n-k} W_0$$

which is simply the binomial expansion for the factors c and $\left[-\lambda B^{-1} + (1+\lambda^2) - \lambda B \right]$.

Thus

$$\gamma_x(B) = \left[c - \lambda B^{-1} + (1+\lambda^2) - \lambda B\right]^n W_0$$

Now $c = (\lambda^2 - \beta)(1-\beta)/\beta$

so

$$\gamma_x(B) = \left[-\lambda B^{-1} + \{(\lambda^2 - \beta)(1-\beta)/\beta\} + 1 + \lambda^2 - \lambda B \right]^n W_0$$

which on simplification gives us

$$\gamma_x(B) = \left[-\lambda B^{-1} + (\beta + \lambda^2/\beta) - \lambda B \right]^n W_0$$

as required.

6.3.2 <u>Comment</u>

Putting $p=1$, $\alpha_1 = 0$ and $W_0 = V$ gives us the proof of the E.W.R. result quoted in section (5.4.2) and further setting $\lambda = 1$ provides the E.W.R. polynomial result given in section (3.5.1).

6.4 Normal Discount Bayesian Model (N.D.B.M.)

A Normal Discount Bayesian Model is defined as

$$Z_t = \underline{u}_t \, \underline{\theta} + \delta_t \qquad ; \delta_t \sim N(\, 0,\ V\,) \tag{6.4.0.1}$$

where Z_t is the derived series for the original observation series Y_t for $t=1,2,\ldots$, such that

$$Z_t = \begin{cases} \sum_{i=0}^{t-1} \psi_i \, \beta^{i/2} \, Y_{t-i} & \text{if } t \le p \\ \sum_{i=0}^{p} \psi_i \, \beta^{i/2} \, Y_{t-i} & \text{if } t > p \end{cases}$$

and (6.4.0.2)

$$\underline{u}_t = \begin{cases} \underline{f} \sum_{i=0}^{t-1} \psi_i \, (\, \beta^{1/2} \, \underline{G}^{-1}\,)^i & \text{if } t \le p \\ \underline{f} \sum_{i=0}^{p} \psi_i \, (\, \beta^{1/2} \, \underline{G}^{-1}\,)^i & \text{if } t > p \end{cases} \tag{6.4.0.3}$$

The AR(p) representation of ε_t is $\phi(B)\varepsilon_t = \psi(B)\varepsilon_t = \delta_t$. Consequently a N.D.B.M formulation based upon Z_t is $\{\underline{u}_t, \underline{G}, V, \beta\}$ which for $t > p$ becomes the constant N.D.B.M. defined over quadruple $\{\, \underline{u}, \underline{G}, V, \beta \,\}$. For example, if $p = 1$, $\underline{f} = (1 \;\; 1)$,

$\underline{G} = \text{diag}(\lambda_1, \lambda_2)$ and $\psi(B) = 1 + \psi_1 B$ where $\psi_0 = 1$ and $\psi_1 = -\phi$ then for some β , Z_t series and $\underline{u}_t$ vectors are derived as follows. At time $t = 1$

$$Z_1 = Y_1 \quad \text{and} \quad \underline{u}_1 = \underline{f} = (1 \quad 1) \;;$$

and at time $t > 1$

$$Z_t = Y_t + \psi_1 \beta^{1/2} Y_{t-1} \quad \text{and} \quad \underline{u}_t = (1+\psi_1 \beta^{1/2}/\lambda_1,\; 1+\psi_1\beta^{1/2}/\lambda_2)$$

$$= \underline{u}$$

a constant vector for all $t > p = 1$. Thus N.D.B.M. at time $t = 1$ becomes constant N.D.B.M when $t > p = 1$. For the derivation of Z_t and $\underline{u}_t$ in the cases of higher degree polynomials, see section (7.5) of chapter seven.

Given a prior $(\underline{\theta}_{t-1} \mid D_{t-1}) \sim N(\underline{m}_{t-1}, \underline{C}_{t-1})$, the posterior $(\underline{\theta}_t \mid D_t) \sim N(\underline{m}_t, \underline{C}_t)$ is found by the recurrence relations as

$$\underline{R}_t = \underline{G}\, \underline{C}_{t-1}\, \underline{G}'/\beta$$

$$\hat{Y}_t = V + \underline{u}_t\, \underline{R}_t\, \underline{u}_t'$$

$$\underline{A}_t = \underline{R}_t\, \underline{u}_t'\, (\hat{Y}_t)^{-1}$$

$$\underline{C}_t = (\, \underline{I} - \underline{A}_t\, \underline{u}_t\,)\underline{R}_t$$

$$e_t = z_t - \underline{u}_t \underline{G}\, \underline{m}_{t-1}$$

$$\underline{m}_t = \underline{G}\, \underline{m}_{t-1} + \underline{A}_t\, e_t \qquad (6.4.0.4)$$

6.4.1 Forecast Function

The k-steps ahead forecast function for the derived series given $D_t = y_t, y_{t-1}, \ldots, y_1$ is

$$E(\, z_{t+k} \mid D_t\,) = \underline{u}_{t+k}\, \underline{G}^k\, \underline{m}_t \qquad (6.4.1.1)$$

and for the original series Y_t, the k-steps ahead forecast function is

$$F_t(k) = \underline{u}_{t+k}\ \underline{G}^k\ \underline{m}_t - \sum_{i=1}^{\ell} \psi_i\ \beta^{i/2}\ x_t(k-i)$$

where $\ell = \text{Min}\ (\ p,\ t-1\ )$

and

$$x_t(k-i) = \begin{cases} y_{t+k-i} & \text{if} \quad k \le i \\ F_t(k-i) & \text{if} \quad k > i \end{cases}$$

(6.4.1.2)

6.4.2 Comment

The derived series Z_t (6.4.0.2) and vector $\underline{u}_t$ (6.4.0.3) are defined using the coefficients of the polynomial $\psi(B)$ up to order p, such that

$$\psi_p(B) = \phi_p(B) = \sum_{i=0}^{p} \psi_i\ B^i = \prod_{i=1}^{p} (\ 1 - \phi_i\ B\ )$$

(6.4.2.1)

For seasonal time series we may consider $\psi(B)$ as

$$\psi_\zeta(B) = \phi_p(B)\ .\ S_s(B) = \sum_{i=0}^{\zeta} \psi_i\ B^i = \prod_{i=1}^{\zeta} (\ 1 - \gamma_i\ B\ )$$

(6.4.2.2)

where $S_s(B)$ is a polynomial in B of degree s for seasonality such that

$$S_s(B) = \sum_{j=0}^{s} S_j B^j = \prod_{j=1}^{s} (1 - \mu_j B) \qquad (6.4.2.3)$$

where the S_j are real but the μ_j occur in complex conjugate pairs. In such a case the series Z_t (6.4.0.2) and the vector $\underline{u}_t$ (6.4.0.3) are redefined, replacing p by $\zeta = p + s$ and using the coefficients ψ_i of the polynomial $\psi_\zeta(B)$.

For practice, a damped form of (6.4.2.3) is recommended. This may be stated as

$$S_s(B) = \prod_{j=1}^{s} (1 - r\mu_j B) \qquad (6.4.2.4)$$

where r is some damping factor lying between 0 and 1. Usually r is very close to 1.

6.5 A Canonical G.E.W.R. Dynamic Linear Model

The model introduced in the previous section is based on the derived series Z_t. For computational purposes it may be desirable to formulate the model in terms of the original series Y_t. It may be required that the high frequency (coloured noise) and the medium frequency (seasonality) is parameterised since they are meaningful and there may be wish to intervene.

A canonical D.L.M. representation of G.E.W.R. is easily obtained as follows:

For any (lxm) vector $\underline{x} = (x_1, \ldots\ldots, x_m)$ define $\underline{J}(\underline{x}) = \{j_{ik}\}$ as the m square matrix such that

$$j_{ik} = \begin{cases} x_i & \text{if } k=i\ ;\ i=1,\ldots\ldots, m \\ 1 & \text{if } k=i+1; i=1,\ldots\ldots,(m-1) \\ 0 & \text{otherwise} \end{cases}$$

Define $\underline{\lambda} = (\lambda_1, \ldots \lambda_n)$, $\beta^{1/2}\underline{\phi} = (\beta^{1/2}\phi_1, \ldots, \beta^{1/2}\phi_p)$

and $\underline{f}_m = (1,0,\ldots,0)$ as a (1xm) row vector with just the unit first element non zero.

A G.E.W.R. Dynamic Linear Model may then be constructed as

$$Y_t = (\underline{f}_p, \underline{0}_n) \tag{6.5.0.1}$$

$$\begin{bmatrix} \underline{\pi} \\ \\ \underline{\rho} \end{bmatrix}_t = \begin{bmatrix} \underline{J}(\beta^{1/2}\underline{\phi} & \underline{0} \\ & \underline{f}_n \\ \underline{0} & \underline{J}(\underline{\lambda}) \end{bmatrix} \begin{bmatrix} \underline{\pi} \\ \\ \underline{\rho} \end{bmatrix}_{t-1} + \begin{bmatrix} \underline{0} \\ \delta \\ \underline{\Delta} \end{bmatrix}_t \tag{6.5.0.2}$$

where at time t,

$$\underline{\pi}_t = (\pi_1,\ldots,\pi_p)'_t \quad \text{and} \quad \underline{\rho}_t = (\rho_1,\ldots,\rho_n)'_t$$

are parameter vectors; and δ_t and $\underline{\Delta}_t$ are white noises having joint distribution as (6.5.0.6).

It is evident that for $t > p$

$$Z_t = \phi(\beta^{1/2} B) Y_t = \underline{f}_n \underline{\rho}_{t-p} + \delta_{t-p+1} \tag{6.5.0.3}$$

$$\underline{\rho}_t = \underline{J}(\lambda)\, \underline{\rho}_{t-1} + \underline{\Delta}_t \tag{6.5.0.4}$$

These equations can be easily written from the above equations (6.5.0.1) and (6.5.0.2) by writing (6.5.0.1) as

$$Y_t = \pi_{1,t} \tag{i}$$

and finding the value of $\pi_{1,t}$ recursively from (6.5.0.2), which is

$$\pi_{1,t} = \{1/\psi(\beta^{1/2} B)\}\{\rho_{1,t-p} + \delta_{t+1-p}\} \tag{ii}$$

where $\psi(\beta^{1/2} B) = \prod_{i=1}^{p} (1 - \beta^{1/2}\phi_i B)$

Substituting (ii) in (i) confirms (6.5.0.3). The expression (6.5.0.4), obviously, directly follows from (6.5.0.2).

The time shift on the parameters and the random variables is immaterial and can be eliminated by re-definition if required. Letting

$$\left(\begin{bmatrix} \underline{\pi} \\ \underline{\rho} \end{bmatrix}_t \middle| D_t\right) \sim N\left[\begin{bmatrix} \hat{\underline{\pi}} \\ \hat{\underline{\rho}} \end{bmatrix}_t, \begin{bmatrix} \underline{\Sigma}_1 & \underline{\Sigma}_2 \\ \underline{\Sigma}_2 & \underline{C} \end{bmatrix}_t\right] \tag{6.5.0.5}$$

and finding

$$\begin{pmatrix} \delta \\ \underline{\alpha} \end{pmatrix}_t \sim N \left[\begin{pmatrix} 0 \\ \underline{0} \end{pmatrix}_t , \begin{bmatrix} V & \underline{0} \\ \underline{0} & (1-\beta)\, \underline{J}(\lambda)\, \underline{C}_{t-1} \underline{J}'(\lambda)/\beta \end{bmatrix} \right]$$

(6.5.0.5)

then for appropriate initialisation, the forecast function is identical to that derived using G.E.W.R. This can be seen by considering the Normal Discount Model given in the previous section and since in terms of similar models discussed in the forthcoming section (6.6), the pair $\{ \underline{f}_n , \underline{J}(\lambda) \}$ is similar to the pair $\{ \underline{f} , \underline{G} \}$.

This canonical model is not usually operationally attractive since the meaning of the parameters is not clear. In practice an appropriate model is easily obtained using the principles given in the following section (6.6).

6.6 Transformation And Similar Models

6.6.1 Transformation

Any observable Linear Dynamic System Model can be transformed from a canonical form ($\underline{f}_n$, $\underline{G}_n$, V_t , $\underline{W}$) to a diagonal form ($\underline{f}_{1,n}$, $\underline{G}_{1,n}$, V_t, $\underline{W}_1$) and vice-versa, through the transformation matrix $\underline{H}_n$. Following Harrison-Akram (1983) we can write

$$\underline{H}_n = \underline{T}_{1,n}^{-1} \underline{T}_n \qquad (6.6.1.1)$$

where

$$\underline{T}_n = \begin{bmatrix} \underline{f}_n \\ \underline{f}_n \underline{G}_n \\ \vdots \\ \underline{f}_n \underline{G}_n^{n-1} \end{bmatrix} \qquad \underline{T}_{1,n} = \begin{bmatrix} \underline{f}_{1,n} \\ \underline{f}_{1,n} \underline{G}_{1,n} \\ \vdots \\ \underline{f}_{1,n} \underline{G}_{1,n}^{n-1} \end{bmatrix}$$

(6.6.1.2)

are each of full rank n . For large n the transfomation matrix $\underline{H}_n$ can be found recursively, without going through the complications of raising the powers of the matrices $\underline{G}_n$ and $\underline{G}_{1,n}$ and finding the inverses of the matrices $\underline{T}_n$ and $\underline{T}_{1,n}$ as

$$\underline{H}_n = \begin{bmatrix} \underline{H}_{n-1} & \underline{B}^{(n)}_{n-1} \\ \underline{0}'_{n-1} & B^{(n)}_{n,n} \end{bmatrix} \qquad (6.6.1.3)$$

for all $n \geq 2$

where
$$\underline{B}^{(n)}_{n-1} = \underline{D}^{(n)}_{n-1} \; \underline{u}^{(n-1)}_{n-1} \qquad ; \; B_0 = 1$$

$$\underline{D}^{(n)}_{n-1} = \text{diag} \{ -(\lambda_n - \lambda_i)^{-1} \} \; ; \; i=1,\ldots\ldots, n-1$$

a (n-1)x(n-1) diagonal matrix. λ_i ($i=1,\ldots\ldots, n-1$) are the non zero eigenvalues of the $\underline{G}_n$ and $\underline{G}_{1,n}$ the similar matrices,

$$\underline{u}^{(n-1)}_{n-1} = (\underline{B}^{(n-1)}_{n-2} \; , \; B^{(n-1)}_{n-1,n-1})'$$

$$B^{(n)}_{n,n} = \prod_{i=1}^{n-1} (\lambda_n - \lambda_i)^{-1} \qquad ; \; B^{(n)}_{1,1} = 0$$

and $\quad u^{(1)}_1 = 1 = H_1$

The inverse of the tranformation matrix $\underline{H}_n$ is

$$\underline{H}^{-1}_n = \begin{bmatrix} \underline{H}^{-1}_{n-1} & \underline{c}^{(n)}_{n-1} \\ \underline{0}'_{n-1} & c^{(n)}_{n,n} \end{bmatrix} \qquad (6.6.1.4)$$

where

$$\underline{c}^{(n)}_{n-1} = (c^{(n)}_1, \ldots\ldots, c^{(n)}_j, \ldots\ldots , c^{(n)}_{n-1})'$$

$$c^{(n)}_j = \begin{cases} 1 & \text{if } j = 1 \\ \prod_{i=1}^{j-1} (\lambda_n - \lambda_i) & \text{if } 2 \le j \le n \end{cases}$$

$$\underline{H}^{-1}_1 = 1 \quad \text{and} \quad c^{(n)}_{n,n} = (B^{(n)}_{n,n})^{-1}.$$

The results (6.6.1.3) and (6.6.1.4) can be easily checked by induction.

Using $\underline{H}_n$ and $\underline{H}^{-1}_n$ we see that

$$\begin{aligned} \underline{f}_{1,n} &= \underline{f}_n \underline{H}^{-1}_n \\ \underline{G}_{1,n} &= \underline{H}_n \underline{G}_n \underline{H}^{-1}_n \\ \underline{W}_{1,n} &= \underline{H}_n \underline{W}_n \underline{H}'_n \end{aligned} \tag{6.6.1.5}$$

i.e we get a diagonal form from the canonical form.

The inverse transformation gives us a canonical form (from diagonal)

$$\begin{aligned} \underline{f}_n &= \underline{f}_{1,n} \underline{H}_n \\ \underline{G}_n &= \underline{H}^{-1}_n \underline{G}_{1,n} \underline{H}_n \\ \underline{W}_n &= \underline{H}^{-1}_n \underline{W}_{1,n} (\underline{H}_n^{-1})'. \end{aligned} \tag{6.6.1.6}$$

6.6.2 Example

Let n=2 $\underline{f} = \underline{f}_2 = (1 \quad 0)$, $\underline{f}_1 = \underline{f}_{1,2} = (1 \quad 1)$

and $\underline{G} = \underline{G}_2 = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{bmatrix}$ and $\underline{G}_1 = \underline{G}_{1,2} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$

$$H_1 = H_1^{-1} = 1 \ , \quad B_0^{(2)} = 1 \ , \quad D_1 = -(\lambda_2-\lambda_1)^{-1} = B_1^{(2)}$$

$$u_1^{(1)} = 1 \quad B_{2,2}^{(2)} = (\lambda_2-\lambda_1)^{-1} \quad \text{and} \quad c_{2,2}^{(2)} = (\lambda_2-\lambda_1)$$

so (6.6.1.3) and (6.6.1.4) reduce to

$$\underline{H}_2 = \begin{bmatrix} 1 & -(\lambda_2-\lambda_1)^{-1} \\ 0 & (\lambda_2-\lambda_1)^{-1} \end{bmatrix}$$

and

$$\underline{H}_2^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & (\lambda_2-\lambda_1) \end{bmatrix}$$

which gives us

$$\underline{f}_1 = \underline{f}\ \underline{H}^{-1} = (1 \quad 0) \begin{bmatrix} 1 & 1 \\ 0 & \lambda_2-\lambda_1 \end{bmatrix} = (1 \quad 1)$$

$$\underline{G}_1 = \underline{H}\ \underline{G}\ \underline{H}^{-1}$$

$$= \begin{bmatrix} 1 & -(\lambda_2-\lambda_1)^{-1} \\ 0 & (\lambda_2-\lambda_1)^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & \lambda_2-\lambda_1 \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

similarly using the inverse transformation we get

$$\underline{f} = \underline{f}_1\ \underline{H} = (1 \quad 1) \begin{bmatrix} 1 & -(\lambda_2-\lambda_1)^{-1} \\ 0 & (\lambda_2-\lambda_1)^{-1} \end{bmatrix} = (1 \quad 0)$$

$$\underline{G} = \underline{H}^{-1}\,\underline{G}_1\,\underline{H}$$

$$= \begin{bmatrix} 1 & 1 \\ 0 & \lambda_2-\lambda_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 & -(\lambda_2-\lambda_1)^{-1} \\ 0 & (\lambda_2-\lambda_1)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{bmatrix}$$

6.6.3 Similar Models

Two observable Dynamic Linear Models defined by the quadruples

$$(\ \underline{f},\ \underline{G},\ V_t,\ W_t\ ) \quad \text{and} \quad (\ \underline{f}_1^*,\ \underline{G}_1,\ V_t,\ \underline{W}_{1,t}\ )$$

are similar if an exact translation of one system to another similar system exists through the transformation matrix $\underline{H}$. In other words if $\underline{G}_1$ is similar to $\underline{G}$ in the sense that there exists a full rank matrix $\underline{H}$ such that

$$\underline{H}\ \underline{G}\ \underline{H}^{-1} = \underline{G}_1 \qquad \text{and} \qquad \underline{f}\ \underline{H}^{-1} = \underline{f}_1^*$$

then it is said that $(\ \underline{f},\ \underline{G}\ )$ is similar to $(\ \underline{f}_1^*,\ \underline{G}_1\ )$. Further if

$$\underline{H}\ \underline{W}_t\underline{H}' = \underline{W}_{1,t} \qquad \text{for all } t$$

then the two Dynamic Linear Models are said to be similar.

If the two observable models are similar then

$$\underline{H} = \underline{T}_1^{-1}\ \underline{T} \qquad \text{and} \qquad \underline{f}_1^* = \underline{f}\ \underline{H}^{-1}.$$

Let the parameterisation of the first model be $\underline{\theta}$ with the prior distribution

$$(\ \underline{\theta}_0\ |\ D_0\ ) \sim N(\ \hat{\underline{\theta}}_0\ ,\ \underline{C}_0\ ).$$

If the parameterisation of the similar model is $\underline{\Gamma}$ then if

$$(\underline{\Gamma}_0 | \; D_0) \sim N(\underline{H} \; \underline{\hat{\theta}}_0 \, , \underline{H} \; \underline{C}_0 \; \underline{H}')$$

the forecast function and all joint forecast distributions are identical for all $t \geq 0$.

The models of a canonical and a diagonal form defined in the example given in the previous sub-section are similar models through the transformation matrix $\underline{H}$.

6.6.4 Comment

The transformation and similar model concept introduced is used to reparameterise the models to a desired form. The recurrence relations developed for the transformation matrices $\underline{H}$ and $\underline{H}^{-1}$, are however for a transformation of a dynamic system from a canonical form to a diagonal form and vice-versa, as generally such transformations are sought. Any other required transformation can be achieved by following the definitions of the transformation matrices $\underline{H}$ and $\underline{H}^{-1}$ i.e.

$$\underline{H} = \underline{T}_1^{-1} \; \underline{T} \qquad \text{and} \qquad \underline{H}^{-1} = \underline{T}^{-1} \; \underline{T}_1$$

For example, the canonical setting for a seasonal model defined over the quadruple ($\underline{f}$, $\underline{G}$, V, $\underline{W}$) with

$$\underline{f} = (1 \quad 0) \quad \text{and} \quad \underline{G} = \begin{bmatrix} e^{i\omega} & 1 \\ 0 & e^{-i\omega} \end{bmatrix} \qquad (6.6.4.1)$$

is undesireable for real valued time series. For operation this model is reparameterised to a similar model defined over the quadruple ($\underline{f}$, $\underline{G}_1$, V, $\underline{W}_1$) with

$$\underline{f} = (1 \quad 0) \quad \text{and} \quad \underline{G}_1 = \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix} \qquad (6.6.4.2)$$

through the transformation matrix

$$\underline{H} = \begin{bmatrix} 1 & 0 \\ i & 1/\sin\omega \end{bmatrix} \tag{6.6.4.3}$$

and its inverse

$$\underline{H}^{-1} = \begin{bmatrix} 1 & 0 \\ -i \sin\omega & \sin\omega \end{bmatrix} \tag{6.6.4.4}$$

$\underline{W}_1$ in this case is of course equal to $\underline{H}\ \underline{W}\ \underline{H}'$.

For more related discussion see Akram-Harrison (1983).

6.7 State Space Representation Of G.E.W.R.

6.7.1 State Space Models

These models are based on the Markov property that implies the independence of the future of the process from its past given the present state. In other words all the information from the past alongwith the present condition of the system which is necessary and sufficient to forecast the future of any stochastic phenomena is summarized into the state of the system, usually called the state vector.

The State Space models lead to recursions which help us to develop the analytical and computational procedures for the estimation of the parameters of a dynamic system.

For related discussion see McGregor (1972), Akaike (1974), Caines-Rissamen (1974), Gardner-Harvey-Philips (1980), Ledolter (1981) and Harrison-Akram (1983).

6.7.2 Construction Procedure

The procedure for finding the E.W.R. type D.L.M's discussed in section (3.5.3) is not always easy but there is a simple way of obtaining a State Space representation for the E.W.R. and the G.E.W.R. ARMA(p,q) cases. However this does not usually provide a desirable parametric form for Bayesian forecasting since the meaning of the parameters is not suited to operation. Further it is often very difficult to transform the parameters to a required set since the necessary transformations are not simple linear ones.

The State Space representation can be constructed as follows:

Any equation of the form

$$\sum_{i=0}^{m} c_i \, y_{t-i} = \sum_{i=0}^{m} b_i \, e_{t-i} \tag{6.7.2.1}$$

where $c_0 = b_0 = 1$ say and where atleast one of c_m and b_m is non zero, can be represented as

$$\begin{aligned} Y_t &= \underline{f}_0 \, \underline{\theta}_t \\ \underline{\theta}_t &= \underline{L} \, \underline{\theta}_{t-1} + \underline{b} \, e_t \end{aligned} \tag{6.7.2.2}$$

where $\underline{f}_0 = (1,0,\ldots,0)$, $\underline{\theta}_t = (\theta_1, \theta_2, \ldots, \theta_{m+1})'$

and $\underline{b} = (b_0, b_1, \ldots, b_m)'$ are (m+1) vectors and

$$\underline{L} = \begin{bmatrix} -\underline{c}_m & \underline{I}_m \\ 0 & \underline{0} \end{bmatrix}$$

is a (m+1) square matrix of rank m,

$\underline{c}_m = (c_1, c_2, \ldots, c_m)'$ is a m vector.

This is easily verified since, defining

$$\theta_{m+2,t} = 0 \qquad \text{for all } t,$$

$$\sum_{i=0}^{m} c_i\, y_{t-i} = \sum_{i=0}^{m} c_i \theta_{1,t-i}$$

$$= \sum_{i=1}^{m+1} (\theta_{i,t-i+1} + c_i \theta_{1,t-i} - \theta_{i+1,t-i}) = \sum_{i=0}^{m} b_i\, e_{t-i} \qquad (6.7.2.3)$$

as from the parameter equation of (6.7.2.1) we see that

$$\theta_{1,t} + c_1 \theta_{1,t-1} - \theta_{2,t-1} = b_0\, e_t$$

$$\theta_{2,t-1} + c_2 \theta_{1,t-2} - \theta_{3,t-2} = b_1\, e_t$$

$$\vdots$$

$$\theta_{m+1,t-m} + c_{m+1} \theta_{1,t-m-1} - \theta_{m+2,t-m-1} = b_m\, e_{t-m}.$$

Hence if $e_t \sim$ Independent $N(\,0, \sigma^2\,)$ a D.L.M. can be written in an extended form with $\underline{\theta}_t$ a $(m+1)$ parametric vector:

$$Y_t = \underline{f}_0\, \underline{\theta}_t$$

$$\underline{\theta}_t = \underline{L}\ \underline{\theta}_{t-1} + \underline{w}_t \qquad (6.7.2.4)$$

where $\underline{w}_t \sim N(\,\underline{0}, \underline{b}\,\underline{b}'\sigma^2\,)$.

In an ARMA(p,q) G.E.W.R. case $\underline{c}_m$ and $\underline{b}$ are obtained from

$$\phi(\,\beta^{1/2}\ B\,) \prod_{i=1}^{n} (\,1-\lambda_i B\,) = \sum_{i=0}^{m} c_i\, B^i$$

$$(6.7.2.5)$$

and $$\eta(\,\beta^{1/2}\ B\,) \prod_{i=1}^{n} (\,1-\beta B/\lambda_i\,) = \sum_{i=0}^{m} b_i\, B^i.$$

Then for $0 < \beta < \mathrm{Min}|\lambda_i^2|$

the $\lim_{t\to\infty} \underline{A}_t = \underline{A} = \underline{b}$ exists, where $\underline{A} = (A_1, \ldots, A_{m+1})$.

This is easy to confirm since at time we can write

$$y_t = \underline{f}_0 \underline{L} \underline{m}_{t-1} + e_t = m_{1,t} + (1-A_1)e_t \qquad (6.7.2.6)$$

as in the limit $t \to \infty$

$$\underline{m}_t = \underline{L} \underline{m}_{t-1} + \underline{A} e_t . \qquad (6.7.2.7)$$

Evaluating $m_{1,t}$ recursively from (6.7.2.7) we see that

$$m_{1,t} = -\sum_{i=1}^{m} c_i m_{1,t-i} + \sum_{i=1}^{m+1} A_i e_{t+1-i} . \qquad (6.7.2.8)$$

Using (6.7.2.6) and (6.7.2.8) we get, on simplification

$$\sum_{i=0}^{m} c_i y_{t-i} = e_t + \sum_{i=1}^{m} \{(1-A_1)c_i + A_{i+1}\} e_{t-i}$$

$$= \sum_{i=0}^{m} b_i e_{t-i} \qquad (6.7.2.9)$$

and clearly $A_1 = 1$ since given y_t, $\theta_{1,t}$ is known to be precisely equal to y_t.

6.8 Variance Learning

For a derived series Z_t , the Normal Discount Bayesian Model (6.4.0.1)

$$Z_t = \underline{u}_t \underline{\theta}_t + \underline{\delta}_t \qquad (6.8.0.1)$$

where $\delta_t \sim N(0, V_t = \phi_t^{-1})$, Z_t and $\underline{u}_t$ are as defined (5.2.2) if V_t is unknown at time t then a method of recurrently estimating it is through the following Bayesian Procedure.

$$V_t = X_t / N_t \qquad (6.8.0.2)$$

$$X_t = \beta_v X_{t-1} + (1- \underline{u}_t \underline{A}_t) e_t^2 \qquad (6.8.0.3)$$

$$N_t = \beta_v N_{t-1} + 1 \qquad (6.8.0.4)$$

where X_t is the sum of squares of the errors associated with the variance V_t, N_t are degrees of freedom, $0 < \beta_v \leq 1$ is a discount factor associated with the variance V_t, e_t and A_t have their usual meaning of one step ahead forecast errors and updating vectors respectively explained in section (6.4).

Basis Of Method

Since the Z_t series is derived from the observation Y_t which are assumed to be normally distributed, the natural posterior distribution for the precision $\S_t$ is Gamma. So letting the prior distributions of $\underline{\theta}_t$, $\S_t$ and $\hat{z}_t$ be

$$(\underline{\theta}_t \mid D_{t-1}, \S_t) \sim N(\hat{\underline{\theta}}_t , \underline{R}_t \S_t^{-1})$$

$$(\S_t \mid D_{t-1}) \sim \Gamma (\beta_v X_{t-1}/2 , \beta_v N_{t-1}/2)$$

$$(\hat{z}_t \mid D_{t-1}, \S_t) \sim N(\hat{y}_t , \hat{Y}_t \S_t^{-1})$$

the posterior distributions of $\underline{\theta}_t$ and $\S_t$ are

$$(\underline{\theta}_t \mid D_t, \S_t) \sim N(\underline{m}_t , \underline{C}_t \S_t^{-1})$$

$$(\S_t \mid D_t) \sim \Gamma (X_t/2 , N_t/2)$$

where

$$\hat{\underline{\theta}}_t = \underline{G} \hat{\underline{\theta}}_{t-1}$$

$$\hat{z}_t = E(z_t)$$

$$\hat{y}_t = E(y_t)$$

and the other symbols have their usual meaning.

Now following Harrison-Johnston (1983) the required recurrence relations are

$$X_t = \beta_v X_{t-1} + (1- \underline{u}_t \underline{A}_t) e_t^2$$

$$N_t = \beta_v N_{t-1} + 1$$

and $$E(\S_t \mid D_t) = V_t^{-1} = N_t / X_t \quad \text{or} \quad V_t = X_t / N_t$$

Initialising X_0 and N_0 such that $V_0 = X_0 / N_0$, along with other prior settings of β_v, $\underline{C}_0$, $\underline{f}$, etc. the system starts updating V_t on the arrival of new information from the observations.

If there is no original information in the system then no contribution to the estimate V_t is made during the first few (say ζ) points. For this period variance learning is not used and instead V_0 is considered for points. In such a case minimum values of X_0 and N_0 are recommended. Moreover $\zeta > n+p+q$ is recommended, where n is the degree of the polynomial required to represent the Low Frequency component or the Trend, p and q are the orders of the Autoregressive and Moving Average processes respectively, required to represent the Coloured Noise or High Frequency component (see Akram-Harrison(1983)

For protection against outliers or to make the learning system robust, the equation (6.8.0.3) is modified as

$$X_t = \beta_v X_{t-1} + (1 - \underline{u}_t \underline{A}_t) d_t \qquad (6.8.0.5)$$

where $d_t = \text{Min.}(e_t^2 , \xi \hat{Y}_t)$ and ξ is a confidence factor

corresponding to a certain level of confidence. For example $\xi = 4$ for 95 % confidence level and $\xi = 6$ for the 99% confidence level.(corresponding to 2σ and 3σ confidence limits respectively).

For variance learning, in order to allow the variance V_t to change slowly, the value of β_v close to one is recommended. For more discussion see Harrison-Johnston (1983).

6.8.1 Comment

The recurrence relations developed for the variance learning of the ARMA-type Coloured Noise processes, based on Bayesian principles improve the performance of the G.E.W.R. models quite significantly.

If $\Psi(B) = 1$ then Z_t and $\underline{u}_t$ reduce to Y_t and $\underline{f}_t$ respectively. In this case the recurrence relations (6.8.0.2) to (6.8.0.5) reduce to the recurrence relations (3.4.1.1) to

(3.4.1.5) given in chapter three. This special case has also been considered by Harrison-Johnston (1983) and Ameen-Harrison (1983). For more related dicussion see West (1982).

7

PRACTICAL ASPECTS OF G.E.W.R.

7.1 Introduction

In this chapter some practical aspects of G.E.W.R. are considered. A Stepwise Identification Procedure (S.I.P) using Average String Lengths (A.S.L.) of the one step ahead forecast residuals is introduced. Theoretical values of the A.S.L. are given in the form of a table for various values of the AR(1) coefficients in order to compare the empirical values of the A.S.L. with these. For visual inspection a graph of the A.S.L. values is presented.

For the sake of writing computer programmes, two forecasting schemes are introduced. These schemes cover most of the E.W.R. and G.E.W.R. type models and operate through the S.I.P. which helps to identify a proper model on the basis of one step ahead forecast errors when the high frequency component can be adequately represented by a low order Autoregressive process.

Three simulation and four real life data analysis are presented. In all cases, the models selected through the S.I.P. generated not only optimum one step ahead forecasts but also quite reasonable long term trends and forecasts as shown in sections (7.4 and 7.5).

7.2 Identification

Various methods and techniques of identifying the order of a given Autoregressive and Moving Average process are cited in the Statistical literature. Among these the Autocorrelation function (A.C.F.) and Partial Autocorrelation function (P.A.C.F.) approach is quite frequently adopted. This approach which helps to identify the order of the ARMA process on the basis of the behaviours of the A.C.F. and P.A.C.F. may be summarised as:

For a p-th order Autoregressive process (AR(p)) the A.C.F. tails off exponentially and P.A.C.F. cuts off after lag p.

For a q-th order Moving Average process (MA(q)) the A.C.F. cuts off after lag q, while at the same lag the P.A.C.F. tails off.

For a mixed p-th order Autoregressive and q-th order Moving Average process (ARMA(p,q)) both the A.C.F. and the P.A.C.F. tail off with a mixture of exponentials and damped sinusoids, after the first (q-p) lags for the A.C.F. and after the first (p-q) lags for the P.A.C.F.

For more discussion see Box-Jenkins (1976). This approach may be used in order to help identify the ARMA type Coloured Noise process and if required, the order of the process may be tested by employing the test procedure given by Godolphin (1980). However, here, a simple approach using Average String Lengths (A.S.L.) is presented as a very quick Stepwise Identification Procedure (S.I.P.) which can be helpful in practice. This approach is developed specially for AR type Coloured Noise processes as in practice our prime interest is in such processes.

7.2.1 Average String Length (A.S.L.)

Average String Length, denoted by A.S.L. is the mean distance (in time units) between the peaks or troughs of the AR(1) type Coloured Noise process ε_t , where

$$\varepsilon_t = \phi\, \varepsilon_{t-1} + \delta_t \qquad (7.2.1.1)$$

$$\delta_t \sim N(\,0,\, \sigma_\delta^2\,) \quad , \quad \varepsilon_t \sim N(\,0,\, \sigma_\varepsilon^2\,)$$

and $\sigma_\varepsilon^2 = \sigma_\delta^2 /(\,1-\phi^2\,)$.

Mathematically , we can define the A.S.L. by

$$\text{A.S.L.} = 1/\{2P(s)\} \qquad (7.2.1.2)$$

where 2P(s) is the limiting probability of a sign change. Thus because of symmetry P(s) is the probability that in the limit $t \to \infty$ a sign associated with a realization of a noise process changes from positive to negative so that at time t

$$P(\varepsilon_{t-1} > 0 > \varepsilon_t) = P(\varepsilon_t > 0 > \varepsilon_{t-1}) = P_t(s)$$

and in the limit $t \to \infty$

$$\begin{aligned} \operatorname{Lim} P_t(s) &= P(s) \\ &= \left(\tan^{-1}\{ -\phi(1-\phi^2)^{-1/2} \} + \pi/2 \right) /2\pi \end{aligned} \tag{7.2.1.3}$$

The expression (7.2.1.3) can be easily proved by considering

$$P_t(s) = P(\varepsilon_{t-1} > 0 > \varepsilon_t)$$

This implies that $\varepsilon_{t-1} > 0$ and $\varepsilon_t < 0$, i.e. $\phi\varepsilon_{t-1} + \delta_t < 0$;

which means that $\varepsilon_{t-1} > 0$ and $\delta_t < -\phi\,\varepsilon_{t-1}$.

In the limit $t \to \infty$, we can write

$$P(s) = \{ (1/\sqrt{2\pi\sigma_\varepsilon^2}) \int_0^\infty e^{-\varepsilon^2/2\sigma_\varepsilon^2} d\varepsilon \} \times \{ (1/\sqrt{2\pi}) \int_{-\infty}^{-\phi\varepsilon/\sigma_\delta} e^{-x^2/2} dx \} \tag{7.2.1.4}$$

Letting $x = \varepsilon u$ and re-arranging the terms we get

$$P(s) = (1/2\pi\sigma_\varepsilon) \int_{-\infty}^{-\phi/\sigma_\delta} du \{ \int_0^\infty e^{-\varepsilon^2(u^2+1/\sigma_\varepsilon^2)/2} \varepsilon\, d\varepsilon \}$$

Further letting $\varepsilon^2(u^2+1/\sigma_\varepsilon^2)/2 = z$ we get

$$P(s) = \frac{1}{2\pi\sigma_\varepsilon} \int_{-\infty}^{-\phi/\sigma_\delta} (u^2 + \sigma_\varepsilon^2)^{-1}\, du \int_0^\infty e^{-z}\, dz.$$

Now $\int_0^\infty e^{-z}\, dz = \Gamma 1 = 1$ and further letting $u = \sigma_\varepsilon^{-1} \tan\alpha$ we get

$$P(s) = \frac{1}{2\pi} \{ \int_{-\pi/2}^{\tan^{-1}(-\phi\sigma_\varepsilon/\sigma_\delta)} \sec^2\alpha\,(1 + \tan^2\alpha)^{-1} \}\, d\alpha$$

$$= \frac{1}{2\pi} \left(\tan^{-1}(-\phi\sigma_\varepsilon/\sigma_\delta) + \pi/2 \right)$$

as $\sec^2\alpha(1 + \tan^2\alpha)^{-1} = 1$.

Substituting $\sigma_\varepsilon = \sigma_\delta/(1-\phi^2)^{1/2}$ in the above expression for $P(s)$, we get

$$P(s) = \left(\tan^{-1} \{ -\phi(1-\phi^2)^{-1/2} \} + \pi/2 \right) / 2\pi$$

the required result (7.2.1.3).

We have considered here the probability that a sign associated with a realization of a noise process changes from positive to negative(from one realization to another realization). Same is the probability that a sign changes from negative to positive. The probability that a sign changes is $2P(s)$ and the expected or the Average String Length (A.S.L), therefore is

A.S.L. = $1/2P(s)$

The Average String Lengths for various values of AR(1) coefficient ϕ have been computed and are given in Appendix A. For visual inspection a graph of A.S.Ls plotted against various values of ϕ is also given in that Appendix. The theoretical values of A.S.Ls so calculated help us to find an approximate value of ϕ , on comparison with the A.S.Ls. found from the one step ahead forecast errors of the time series under study. The procedure adopted , i.e. a Stepwise Identification Procedure (S.I.P.) is discussed in sub-section (7.2.3).

7.2.2 Examples

7.2.2.1 Example 1

500 realizations of an AR(1) Coloured Noise process ε_t with coefficient $\phi = -0.9$ are generated so that

$$\varepsilon_t = -0.9\,\varepsilon_{t-1} + \delta_t \qquad \text{for } t = 1,\ldots,500$$

where $\delta_t \sim IN(0, 10000)$

and are given in Appendix C.

The A.S.L. calculated for the generated data is found approximately equal to 1.15. This value closely tallies with the theoretical value of A.S.L. = 1.17 , given in Appendix A, corresponoding to the value of AR(1) coefficient $\phi = -0.9$. According to test of significance (7.2.3), there does not seem to be significant difference between the empirical and the theoretical values of the A.S.Ls at 1% level of significance. This information suggests AR(1) type G.E.W.R. model with AR(1) coefficient ϕ approximately equal to -0.9. The G.E.W.R model so selected, quite likely, will filter out the coloured noise and evaluate optimum one step ahead forecasts. One step ahead forecast errors in this case will form a white noise sequence with zero mean and variance σ^2 as t becomes large. The A.S.L. value in this case would not be significantly different from 2, which is the theoretical value of the A.S.L. corresponding to $\phi = 0.0$, i.e. the white noise case.

7.2.2.2 Example 2

In this example, 500 realizations of an AR(1) coloured noise process ε_t with coefficient $\phi = 0.5$ are generated so that

$$\varepsilon_t = 0.5\ \varepsilon_{t-1} + \delta_t \qquad \text{for } t = 1, \ldots, 500$$

where $\delta_t \sim IN(0, 10000)$

and are given in Appendix E. The A.S.L. calculated in this case is approximately equal to 3.09. The theoretical value of A.S.L. (from the Appendix A) for $\phi = 0.5$ is 3.00. Applying test of significance (7.2.3), there does not seem to be any significant difference between the empirical and the theoretical values of the A.S.Ls. at 1% level of significance. This information suggests an AR(1) type G.E.W.R. model with AR(1) coefficient $\phi \simeq 0.5$. The G.E.W.R. model so selected, as in the case of previous example, is likely to yield optimum one step ahead forecasts with one step ahead forecast errors forming a white noise sequence, yielding an A.S.L. value not significantly different from 2.

Comment

The theoretical values of the A.S.Ls. given in Appendix A are limiting values, computed by using the limiting expression (7.2.1.2). The Average String Lengths calculated from the one step ahead forecast errors incurred by some appropriate forecasting model or from some coloured noise process vary over time. These empirical A.S.Ls. values tend to approach the theoretical A.S.Ls. values as t becomes large, i.e. as $t \to \infty$

$$\underset{\text{(empirical)}}{\text{A.S.Ls}} \to \underset{\text{(theoretical)}}{\text{A.S.Ls.}}$$

This follows from frequency distribution of probability.

7.2.3 Testing The Whiteness Of The Residual Errors

There are various methods which can be employed to check whether a sequence of residuals or forecasting errors is a Gaussian White Noise sequence or not. Here, a very simple and straightforward procedure is given, which quite effectively tests the Whiteness of the residual errors. In this method, we test whether an empirical value of the Average String Length (A.S.L.) evaluated from the residual or the one step ahead forecast errors incurred by an E.W.R. type or a G.E.W.R. type model is significantly different from 2 or not; where 2 is the theoretical value of A.S.L. (given in Appendix A) when the AR(1) coefficient $\phi = 0$.

Let X be a random variable (r.v.) representing the number of sign changes in a White Noise sequence of $N + 1$ observations, then given the first sign (positive or negative) the r.v. X is binomially distributed with mean $N/2$ and variance $N/4$, i.e.

$$X \sim B(N/2 , N/4).$$

For large number of realizations, i.e. $N \geq 25$, the confidence interval for the number of sign changes is given as:

$$\{ N/2 \pm Z_{\alpha/2} \sqrt{N}/2 \} \qquad (7.2.3.1)$$

where for any given value of α , the level of significance, $Z_{\alpha/2}$ is the value of the standard normal variate Z corresponding to $\alpha/2$. For the parent A.S.L. the confidence interval is given as

$$\{ 2(N+1)/(N+Z_{\alpha/2} \sqrt{N}), 2(N+1)/(N-Z_{\alpha/2}\sqrt{N}) \} \qquad (7.2.3.2)$$

In the limit $N \to \infty$, both the lower and the upper values of the confidence interval (7.2.3.2) approach 2, which is the theoretical value of the A.S.L. corresponding to the AR(1) coefficient $\phi = 0$.

In order to test the significance of the empirical value of the A.S.L. we formulate the Null Hypothesis (H_0) as

H_0 : A.S.L. = 2 (i.e. the noise is white)

along with the one-sided Alternative Hypotheses (H_1 and H_2) as

H_1 : A.S.L. < 2 (i.e. the noise is coloured with some negative value of AR(1) coefficient)

and H_2 : A.S.L. > 2 (i.e. the noise is coloured with some positive value of AR(1) coefficient)

The decision to use H_1 or H_2 along with H_0 depends upon the empirical value of A.S.L. and our interest in detecting or identifying the type of coloured noise.

The critical regions (C.Rs.) are defined as

i) C.R.1 : region $< 2(N+1)/(N+Z_\alpha \sqrt{N})$

ii) C.R.2 : region $> 2(N+1)/(N-Z_\alpha \sqrt{N})$ (7.2.3.3)

The Null Hypothesis H_0 is accepted at certain value of α, the level of significance, if the empirical value of A.S.L. lies outside the critical regions, otherwise reject it.

If the empirical value of A.S.L. falls in the critical region C.R.1, we reject the Null Hypothesis H_0 and accept the Alternative Hypothesis H_1.

If the empirical value of A.S.L. falls in the critical region C.R.2, we reject H_0 and accept H_2.

In both of the Null Hypothesis rejection cases, we consult the table (Appendix A) of theoretical values of the A.S.Ls. in order to find an appropriate value of the AR(1) coefficient of the coloured noise.

7.2.4 Stepwise Identification Procedure (S.I.P.)

In practice the noise process for a G.E.W.R. application is well represented by an AR(p) process of order p = 1 or p = 2. Yule-Walker equations (Yule(1927) and Walker(1931)), Autocorrelation Function (A.C.F.) and Partial Autocorrelation Function (P.A.C.F.) approaches may be used for identification. However, here a simple approach using Average String Lengths (A.S.Ls.) is presented as a very quick Stepwise Identification Procedure (S.I.P.) which can be helpful in practice.

Step 1

An E.W.R. type model (Constant Dynamic Linear Model, D.E.W.R., etc., as discussed in previous chapters) is applied with the settings $\underline{f}$, $\underline{G}_L$, $\underline{W}$, β , β_V (a variance discount factor), etc. as required and the prior

$$\underline{\theta}_0 \sim N(\underline{m}_0, \underline{C}_0) , V_0 \ \& \ N_0 .$$

The discount factor β is selected close to one, representing the very slow change in the low frequency component.The Average String Length (A.S.L.) of the one step ahead forecast errors incurred by the fitted model is obtained and compared with the theoretical value of the A.S.L. given in the Appendix A, corresponding to AR(1) coefficient ϕ = 0.00, which is 2 in this case. If the empirical value of the A.S.L. is not significantly different from two, then this suggests that the one step ahead forecast errors form a sequence (as t becomes large) from a white noise process with mean zero and some variance, say σ^2 . This means that the time series under study is not driven by a coloured noise process, but instead bears a sequence of a white noise process.

The procedure adopted for forecasting is according to the forecasting schemes given in section (7.3). An appropriate scheme is considered according to the forecasting model selected in order to analyse a time series.

If the empirical value of A.S.L. is significantly different from two, we proceed to step 2.

Step 2

This step is carried out if either the empirical value of A.S.L. is significantly less than two or significantly greater than two.

An empirical value of A.S.L. significantly less than two suggests that the AR(1) coefficient ϕ is negative. In such a case an approximate value of ϕ is found from the Appendix A. For example, A.S.L. $\cong$ 1 suggests $\phi \cong -0.9$.

An empirical value of A.S.L. significantly greater than two suggests that the coefficient ϕ is positive, i.e. $\phi > 0$. For example A.S.L. $\cong$ 4 suggests $\phi \cong 0.7$.

In any of the above situations we adopt a G.E.W.R. type model with an AR(1) high frequency component and the prior as for the E.W.R. type model (step 1). One step ahead forecast errors incurred by the G.E.W.R. type model are again inspected and A.S.L. computed for these residual errors. If the empirical value of A.S.L. does not differ significantly from two, we stop and consider the adopted model as an appropriate model. If the empirical value of A.S.L. evaluated from the residual errors differ significantly from two, we proceed to step 3.

Step 3

We proceed to this step from step 2 if we fail to achieve the residual errors or one step ahead forecast errors as a sequence of a white noise process using an AR(1) type G.E.W.R. model. Here, again there are two possibilities for the empirical value of the A.S.L. evaluated after processing step 2. One is that the A.S.L. is significantly less than two and the other that it is significantly greater than two. In any case we proceed to a G.E.W.R. type model with an AR(2) high frequency component.

For this G.E.W.R. type model the AR(2) coefficients ϕ_1 and ϕ_2 may approximately be selected as follows. The AR(1) coefficient ϕ selected during the course of step 2 is retained as ϕ_1 and the second coefficient ϕ_2 is selected approximately on the suggestion of the empirical value of A.S.L. evaluated from the residual errors or one step ahead forecast errors incurred by the AR(1) type G.E.W.R. model. The stated method of finding ϕ_1 and ϕ_2 is not unique. Some other method may be employed, but in practice, this method is found to be quite useful (for example see the case studies (7.5)).

In most practical situations a G.E.W.R. type model with AR(2) coloured noise component is sufficient. However, if one step ahead forecast errors incurred by the AR(2) type G.E.W.R. model do not form a sequence of a white noise process, i.e. the empirical value of the A.S.L. is significantly different from 2, then we proceed to G.E.W.R. type model with AR(3) coloured noise component and repeat the analysis process following the previous steps. We go on cycling by increasing the order of the coloured noise component (say up to order p) associated with the G.E.W.R. type models, until we get the one step ahead forecast errors forming a sequence of a white noise process (A.S.L., not significantly different from 2).

In order to test the significance of the empirical values of the A.S.L. we follow the test procedure given in the previous section. Stepwise identification is carried out following the forecasting schemes given in the next section.

7.3 Forecasting Schemes

In order to identify a proper forecasting model and implement the computer programmes for the k-steps ahead ($k \geq 1$) forecasts, two layouts of the forecasting schemes, known as Forecasting Schematas (F.Sc.) are presented.

Forecasting Schemata 1 (F.S.1) is designed for the E.W.R. type and the G.E.W.R. type Dynamic Linear Models. In this scheme low, medium and high frequency components are accommodated within the transition matrix $\underline{G}$ such that $\underline{G} = \mathrm{diag}(\underline{G}_h, \underline{G}_s, \underline{G}_\ell)$

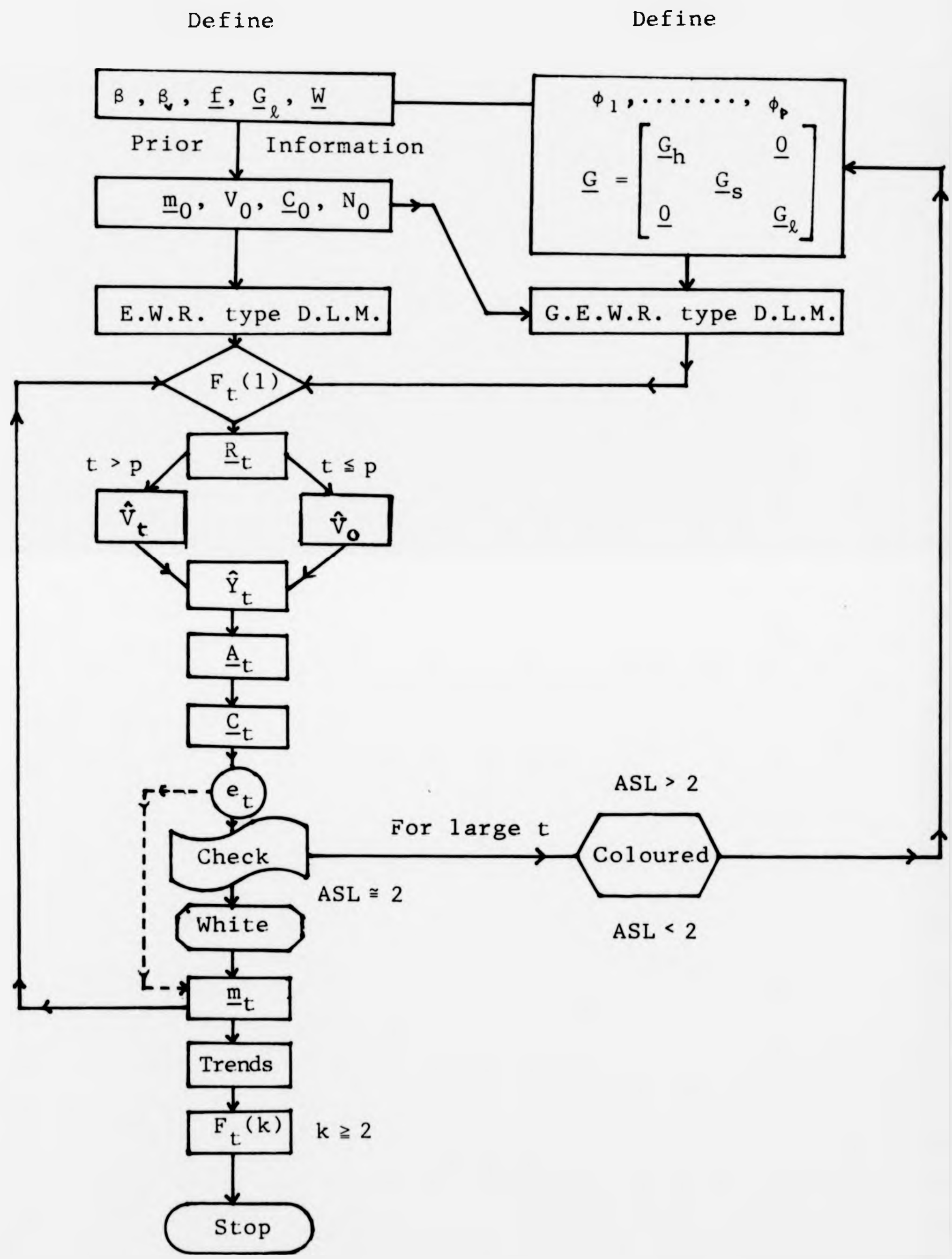
where $\underline{G}_h$, $\underline{G}_s$, $\underline{G}_\ell$ are the sub-matrices for the high (Coloured Noise), medium (Seasonal) and the low (Trend) frequencies. The dynamic system may be of any form (canonical, diagonal, etc.). The $\underline{W}$ matrix is set in a canonical form as explained in the previous chapters and then transformed to any required form following the transformation techniques given in chapter six.

Forecasting Schemata 2 (F.S.2) is designed for the E.W.R. type and the G.E.W.R. type D.E.W.R. and the Normal Discount Bayesian Models, where high and medium frequencies are accommodated in $\underline{u}_t$ vectors and a Z_t series is generated from the original time series Y_t in order to find the k steps ahead forecasts and the trends. Low frequency is translated through the $\underline{G}$ matrix.

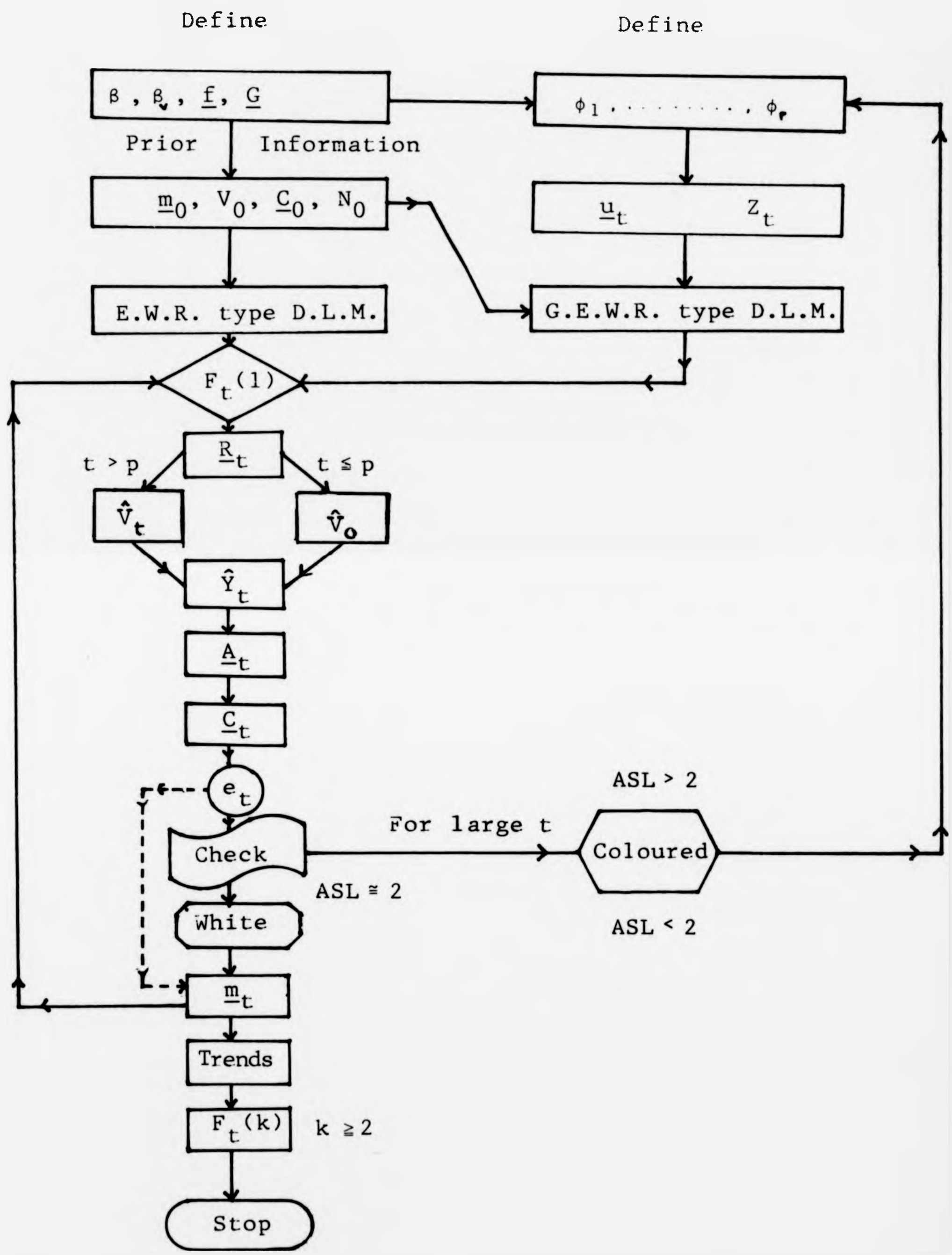
In both the forecasting schemes, provision is made for on-line variance learning and checking the nature of the forecasting or residual errors (i.e. whether the residual errors form a sequence of a white noise process or a coloured noise process). Both the schemes are equally valid for one step ahead and k ($k \geq 2$) steps ahead forecasts and trends. The Forecasting Schematas not only provide the insight in to dynamic systems but also help us to write the computer algorithms and programmes.

Both the forecasting schemes have been tested on the simulated and real life data sets, but here in this dissertation F.S.1 for the simulated data sets and F.S.2 for the real life data sets are presented.

FORECASTING SCHEMATA 1



FORECASTING SCHEMATA 2



7.4 Simulation

In order to demonstrate the performance of G.E.W.R. methodology, it is applied to various simulated stochastic data sets which comprise low frequency (Trend) and high frequency (Coloured Noise) components. The linear growth type low frequency is generated by considering

$$\underline{m}_t = \underline{G}_\ell \, \underline{m}_{t-1} + \underline{A} \, \xi_t \qquad (7.4.1)$$

where

$$\underline{G}_\ell = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad , \quad \underline{A} = \begin{pmatrix} 1-\beta^2 \\ (1-\beta)^2 \end{pmatrix} ,$$

ξ_t is a White Noise process with mean zero and variance 1 and discount factor $\beta = 0.99$. The initial prior value is

$$\underline{m}_0 = (100 \quad 1)'$$

The high frequency component which is superimposed on the linear growth low frequency is generated using an AR(1) Coloured Noise model

$$\varepsilon_t = \phi \, \varepsilon_{t-1} + \delta_t$$

where $\delta_t \sim IN(0, 10000)$. and $| \phi | < 1$.

Cases for various values of ϕ are studied, but for the three cases presented here the values of ϕ considered are:

i) Case 1 : $\phi = - 0.9$

ii) Case 2 : $\phi = 0.3$

iii) Case 3 : $\phi = 0.5$

The generated data sets, each consisting of 500 realizations are given in Appendices B, C and D.

In each case first an E.W.R. type Dynamic Linear Model (3.5.1.2) of the following form is fitted:

$$Y_t = \underline{f}\,\underline{\theta}_t + v_t \qquad ; v_t \sim N(0, V)$$

$$\underline{\theta}_t = \underline{G}_\ell\,\underline{\theta}_{t-1} + \underline{w}_t \qquad ; w_t \sim N(0, W_\ell) \qquad (7.4.2)$$

where $\underline{f} = (1 \quad 0)$, $\underline{G}_\ell = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$,

$$V = 10000 \quad , \quad \underline{W}_\ell = \mathrm{diag}(W_{\ell 1}, W_{\ell 2})$$

$$W_{\ell j} = \binom{2}{j} \{ (1 - \beta)^2 / \beta \}^j \, V \quad \text{for } j=1,2 \qquad (7.4.3)$$

and $\beta = 0.995$

is applied. The prior considered is

$$\underline{m}_0 = (100 \quad 1) \quad \text{and} \quad \underline{C}_0 = \mathrm{diag}(10000, 100).$$

The forecasting scheme is implemented according to the Forecasting Schemata 1. In all the three cases, Average String Lengths (A.S.Ls.) of the residuals or one step ahead forecast errors incurred by the above model are evaluated on the computer (within the computer programme). The A.S.L values found are 1.15, 2.48 and 3.07 for case 1, case 2 and case 3 respectively.

The empirical values of A.S.L. evaluated are tested for significance by employing the testing of Whiteness procedure given in section (7.2.3) at the level of significance $\alpha = 0.05$. In all cases the empirical values of A.S.L. are found to be significantly different from 2. This means that the model used is not suitable for the data sets under study, as the selected model should be capable of evaluating the forecasts such that the residual errors or one step ahead forecast errors form a sequence of a White Noise process. We need to proceed further, i.e. to G.E.W.R. type models, following the Stepwise Identifican Procedure (S.I.P.) described in section (7.2.4).

In case 1, the empirical value of A.S.L. falls in the critical region 1 (C.R.1 : region <1.84 for N=499 and $\alpha = 0.05$).

In cases 2 and 3, the empirical values of A.S.L. fall in the critical region 2 (C.R.2 : region >2.2 for N=499 and $\alpha = 0.05$).

Comparing the empirical values of A.S.L. with the theoretical values of A.S.L. given in Appendix A, we find the estimates of the AR(1) coefficient ϕ as approximately -0.91, 0.30 and 0.52 in cases 1, 2 and 3 respectively.

Following S.I.P. we proceed to first order Autoregressive G.E.W.R. type Constant Dynamic Linear Model similar to the model (6.2.1.1), where in each case for the high frequency component the approximately estimated values of ϕ are used. The general form of the model used is

$$Y_t = \underline{f}\,\underline{\theta}_t$$

$$\underline{\theta}_t = \underline{G}\;\underline{\theta}_{t-1} + \underline{w}_t \qquad ; \underline{w}_t \sim N(\,0,\,\underline{W}_1\,) \tag{7.4.4}$$

where $\underline{f} = (1 \quad 0 \quad 1)$, $\underline{G} = \begin{bmatrix} \underline{G}_\ell & \underline{0} \\ \underline{0} & a \end{bmatrix}$, $\underline{G}_\ell = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

$\beta = 0.995$ and $a = \beta^{1/2}\,\phi$

The matrix $\underline{W}_1 = \underline{H}\,\underline{W}\,\underline{H}'$ is found following the procedure for obtaining the similar models, explained in section(6.6), where

$$\underline{W} = \text{diag}(\,V,\,W_{\ell 1},\,W_{\ell 2}\,)$$

is a matrix for the canonical form of a dynamic system, $W_{\ell 1}$ and $W_{\ell 2}$ are as defined earlier, $\underline{H}$ is a transformation matrix (6.6) that transforms the $\underline{W}$ matrix (canonical form) to a dynamic system defined by the above model. Transformation is carried out on the computer (within the computer programme). The prior infomation provided to the model is

$$\underline{m}_0 = (\,100 \quad 1 \quad 0\,)' \quad \text{and} \quad \underline{C}_0 = \text{diag}(\,10000,100,10\,)$$

One step ahead forecasts and residuals found by using the recurrence relations given in section (6.2.1), replacing $\underline{J}$ by $\underline{G}$. Average string lengths of the residuals incurred by the model (7.4.4), for $\hat{\phi} = -0.91$ in case 1, $\hat{\phi} = 0.30$ in case 2 and $\hat{\phi} = 0.52$ in case 3 are evaluated and tested for the Whiteness of the residuals, again following the test of significance procedure (7.2.3). In all three cases the empirical values of A.S.L. are found to be not significantly different from 2 at the level of significance $\alpha = 0.05$. This means that in all the three cases the models with the estimated values of the Autoregressive coefficient ϕ , identified by the S.I.P. behaved well and yielded residuals or one step ahead forecast errors forming a sequence of a White Noise process. In all cases one step ahead forecasts found are also optimum in the Minimum Mean Square Error (M.M.S.E.) sense.

After making sure (on the basis of test of significance (7.2.3)) that the one step ahead forecast errors or residuals are free from the coloured noise, long term (k-steps ahead, $k \geq 2$) forecasts are found from the forecast function

$$F_t(k) = \underline{f}\ \underline{G}^k\ \underline{m}_t$$

following the Forecasting Schemata 1.

In case 1 ($\phi = -0.9$), short and long term (up to 60 steps ahead) forecasts are found and are shown in figures 1b to 1h. Figure 1a shows the graph of the data which comprise low frequency and high frequency ($\phi = -0.9$). Figure 1b shows the one step ahead forecasts. Long term forecasts (for k=5, 10, 20, 30, 40, 50 and 60) are shown in Figures 1c to 1g. The k-steps ahead forecasts seem to be quite reasonable and the successive forecasts show that in the long run the low frequency (Trend) is well protected from the high frequency (Coloured Noise).

In case 2 ($\phi = 0.3$), up to 10 steps ahead forecasts are found and are shown in Figures 2b to 2d. Figure 2a shows the simulated data series comprising of low frequency and high frequency ($\phi = 0.3$). One step ahead forecasts are shown in Figure 2b. Long term forecasts, 5 steps and 10 steps ahead are shown in Figures 2c and 2d respectively. Figure 2e shows the residuals incurred by the E.W.R. type Dynamic Linear Model (7.4.2). The residuals gave us a value of the A.S.L. equal to 2.48. In Figure 2f, the low frequency component generated by the model (7.4.1) is shown. This low frequency is the same for all the three cases. Long term (10 steps ahead) forecasts (Fig. 2d) seem to be approaching the underlying trend or low frequency (Fig. 2f). Thus in the long run the low frequency is well protected from the high frequency in our joint modelling scheme. This is true in all cases.

In case 3 ($\phi = 0.5$), as usual, up to 20 steps ahead forecasts are found but up to 10 steps ahead forecasts are presented (Figs. 3b to 3d) as 10 steps ahead forecasts clearly show that the low frequency is well protected from the high frequency and it is true for all forecasts higher than 10 steps ahead. In Figure 3a the actual data series is shown which is comprised of low frequency and high frequency ($\phi = 0.5$) component. Figure 3b shows the one step ahead forecasts. Figures 3c and 3d show the 5 steps ahead and 10 steps ahead forecasts respectively.

AR(1) Coloured Noise generated with mean zero, variance 10,000 and $\phi = -0.9$; superimposed on a Linear Growth Level with $\beta = 0.995$.

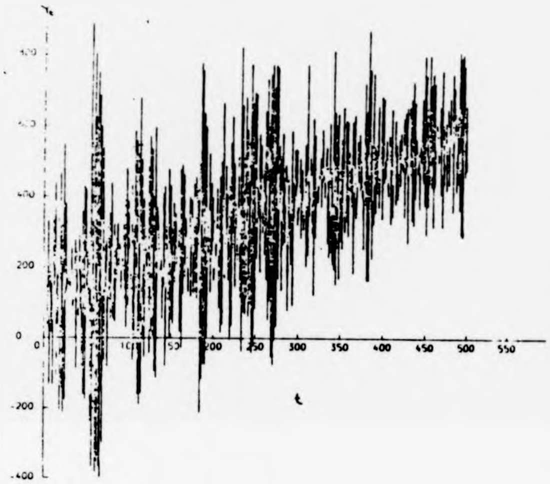


Fig. 1a

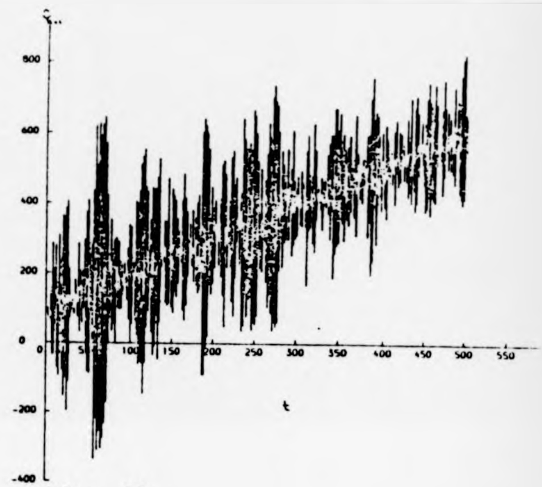


Fig. 1b

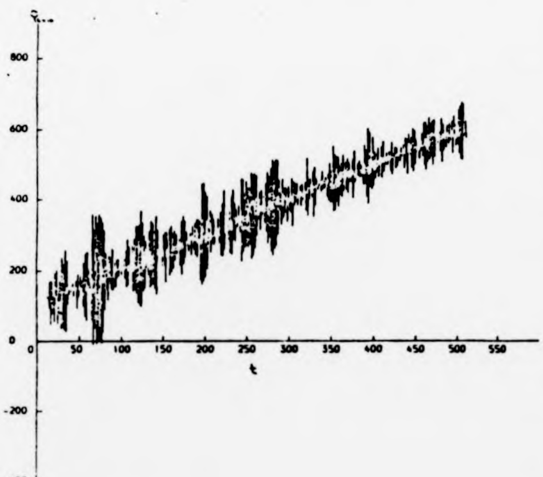


Fig. 1c

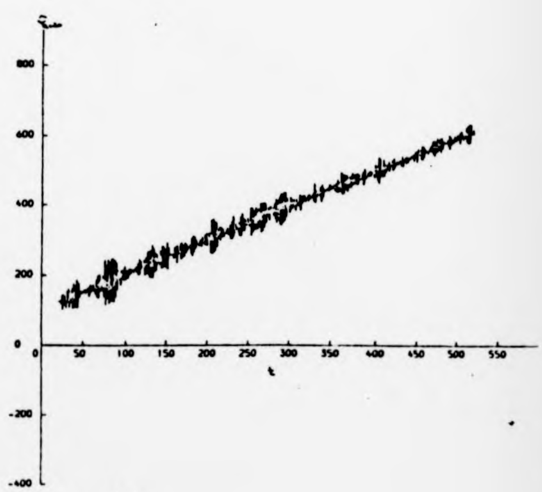


Fig. 1d

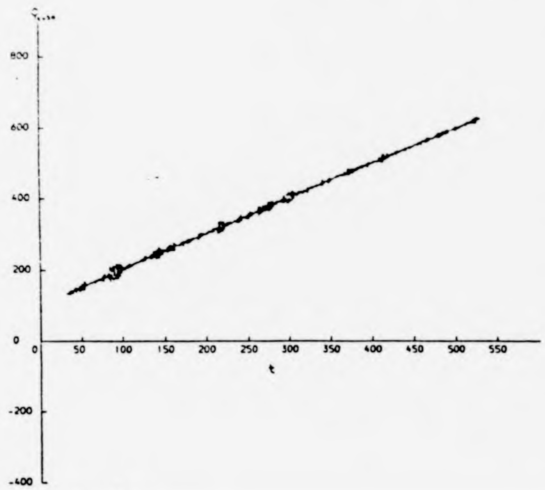


Fig. 1e

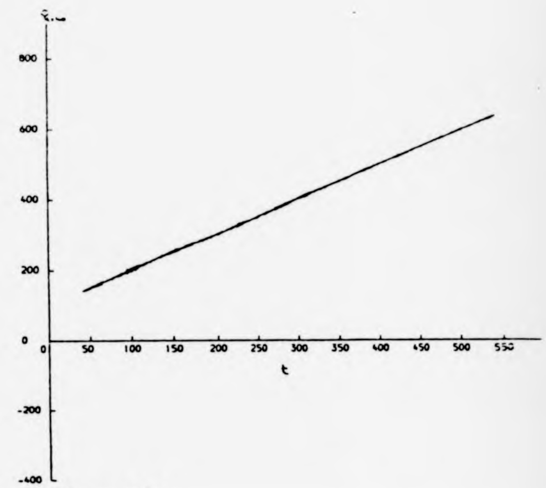


Fig. 1f

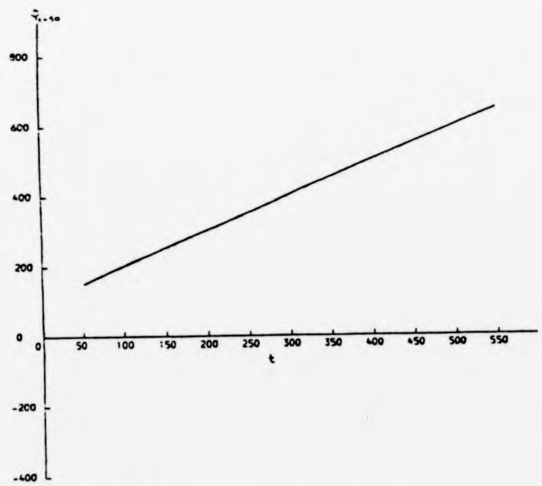


Fig. 1g

AR(1) Coloured Noise generated with mean zero, variance 10,000 and ϕ =0.3; superimposed on a Linear Growth Level with β =0.995.

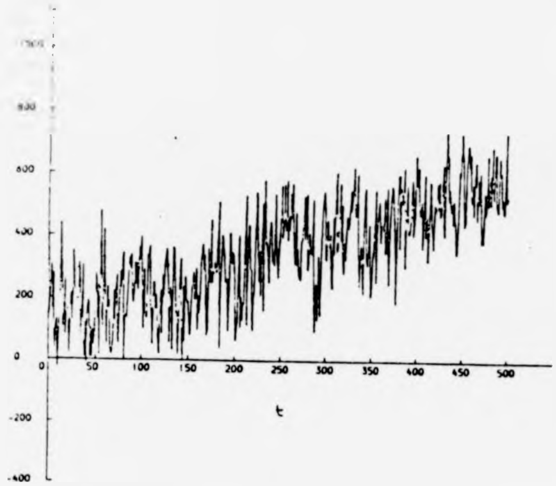


Fig. 2a

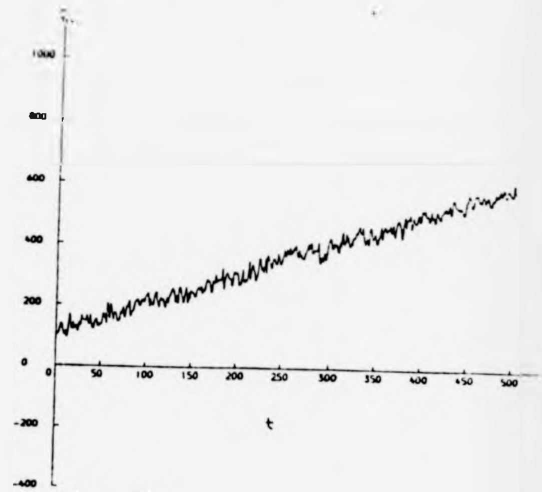


Fig. 2b

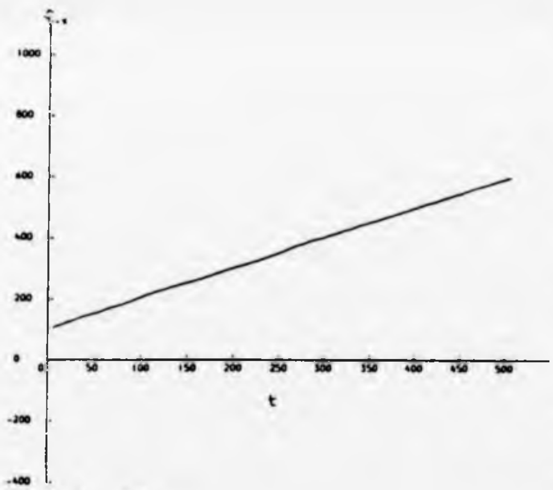


Fig. 2c

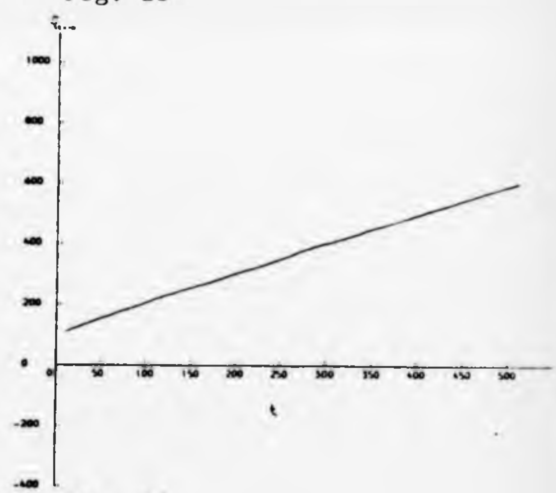


Fig. 2d

Residual Errors incurred by the E.W.R. type model (7.4.2) ($\phi = 0.3$)

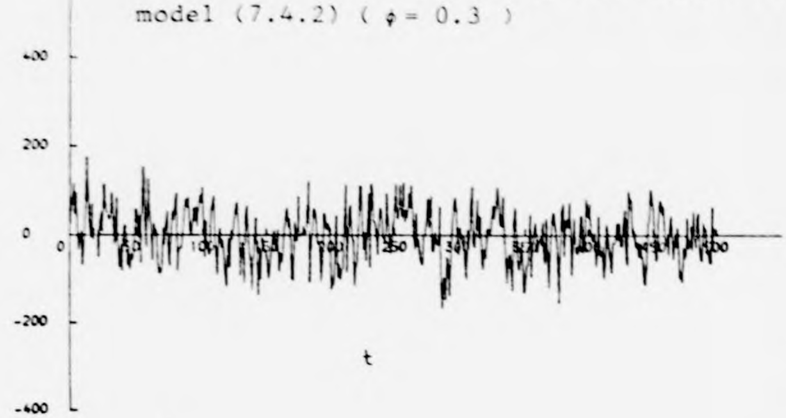


Fig. 2e

Low frequency (Trend) generated by Linear Growth Model (7.4.1)

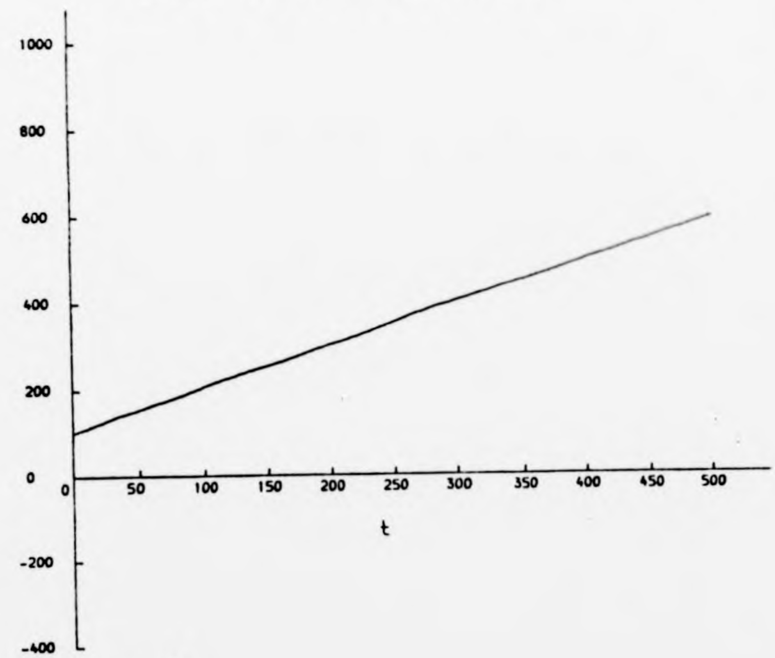


Fig. 2f

Residual Errors incurred by the E.W.R. type model (7.4.2) ($\phi = 0.3$)

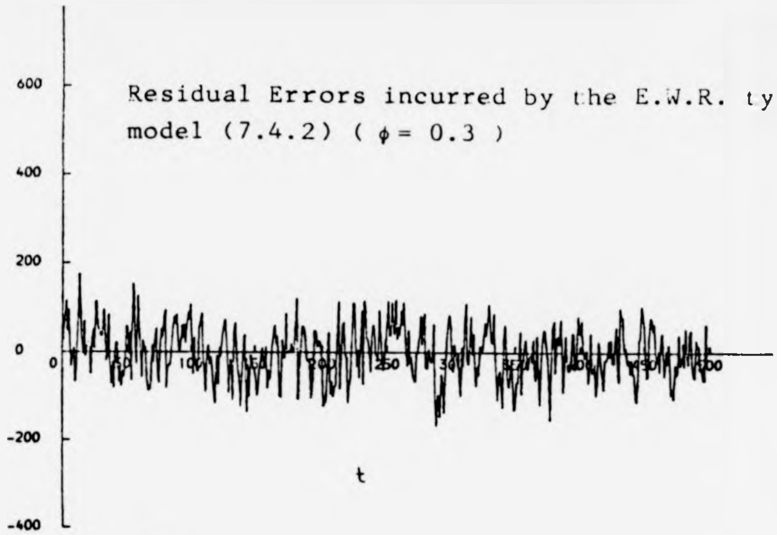


Fig. 2e

Low frequency (Trend) generated by Linear Growth Model (7.4.1)

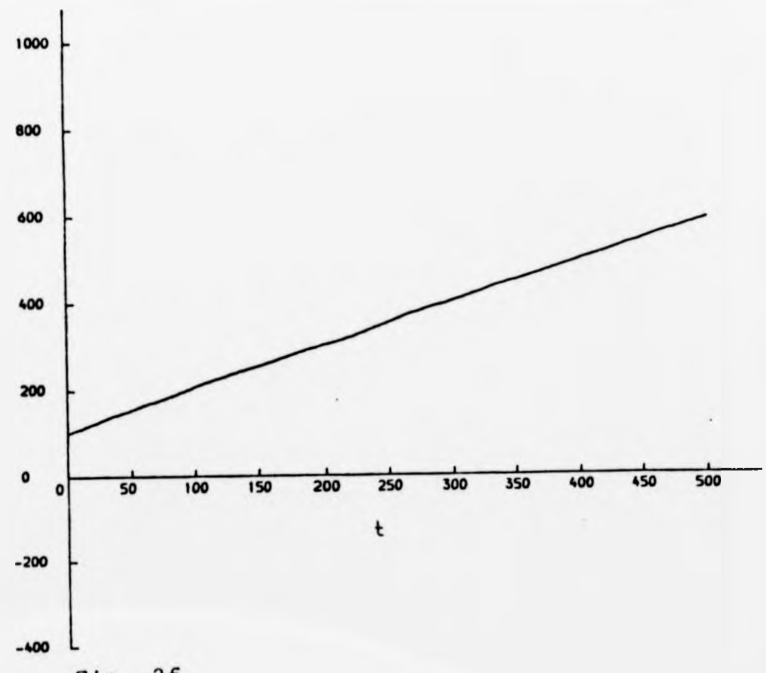


Fig. 2f

AR(1) Coloured Noise generated with mean zero, variance 10,000 and $\phi = 0.5$; superimposed on a Linear Growth Level with $\beta = 0.995$

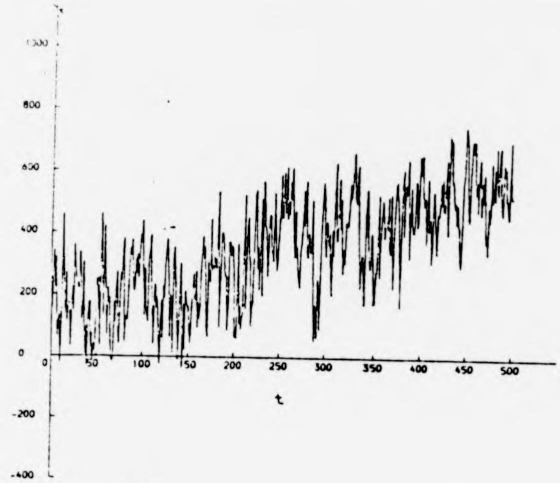


Fig. 3a

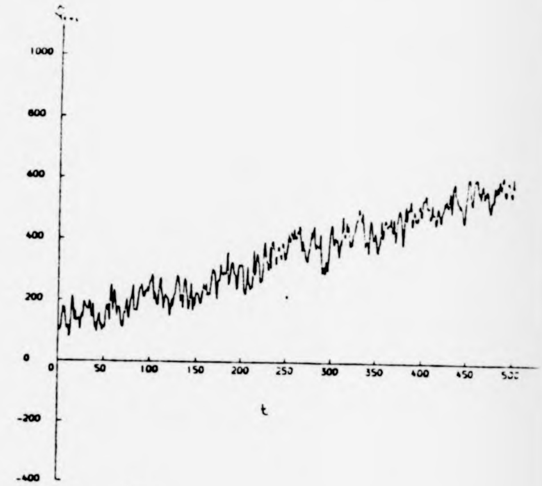


Fig. 3b

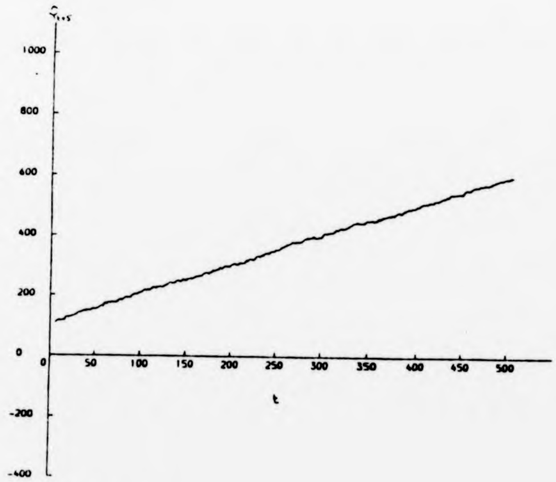


Fig. 3c

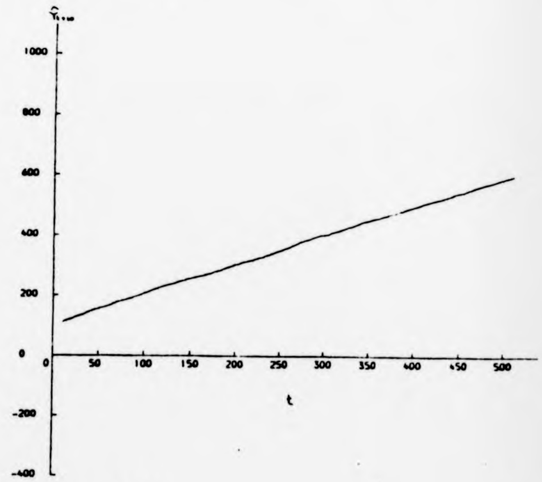


Fig. 3d

7.5 Case Studies

Four cases of real life data, two seasonal and two non-seasonal are studied. The cases are:

i) Case 1 : Availability of Electricity in Great Britain from 1901 to 1980. Annual data (given in Appendix G).

ii) Case 2 : U.K. Chemical Industries Indices from 1952 to 1982. Quarterly data (given in Appendix H).

iii) Case 3 : Population of U.S.A. from 1901 to 1981. Annual data (given in Appendix I)

iv) Case 4 : Austrian Disposable Personal Income from 1954 to 1979. Quarterly data (given in Appendix J).

In all cases short to long term forecasts (up to 20 steps ahead) are found along with the trends by using G.E.W.R. methodology, following the Stepwise identification procedure (S.I.P.) and the forecasting scheme described by the Forecasting Schemata 2. The joint modelling scheme for low frequency (Trend), medium frequency (Seasonal) and high frequency (Coloured Noise) is adopted.

In all cases the Normal Discount Bayesian Model (N.D.B.M.) (6.4.0.1)

$$Z_t = \underline{u}_t \, \underline{\theta} + \delta_t \quad ; \delta_t \sim N(0, V_t) \qquad (7.5.0.1)$$

is applied in the diagonal form with

$$\underline{f} = (1 \quad 1) \quad \text{and} \quad \underline{G} = \text{diag}(1, \lambda).$$

For the low frequency characterized by ($\underline{f}$, $\underline{G}$, β) a log analogue of the Gompertz function

$$y_t = a\, b^{\rho^t} \qquad ; a > 1 \, , \quad 0 < b,\rho < 1 \qquad (7.5.0.2)$$

i.e.

$$\text{Log } y_t = \text{Log } a + (\text{ Log } b) \rho^t$$

$$\bar{y}_t = A + B \rho^t \qquad \text{is used.}$$

The choice of Gompertz function is made owing to the fact that it adequately represents many economic, social and industrial phenomena that typically follow ' S ' shape curves close to the Gompertz form and tend to an asymptote or saturation level. The parameters A, B, ρ are evaluated following the procedure described by Harrison-Pearce (1972) and Stoodley (1980). In order to establish a link between the Gompertz function and our dynamic model, the value λ of the transition matrix $\underline{G}$ is estimated by comparing the estimator

$$\hat{\bar{y}}_t = \hat{A} + \hat{B} \hat{\rho}^t \tag{7.5.0.3}$$

with the k steps ahead forecast function (for trend)

$$F_t(k) = m_t + \lambda^k b_t \tag{7.5.0.4}$$

This gives us an approximate value of λ that enables the dynamic system to transit through time in a Gompertz way.

The series Z_t is derived from the original time series Y_t and the vector $\underline{u}_t$ is found, following the definitions (6.4.0.2) & (6.4.0.3) and considering

$$\psi_p(B) = \phi_p(B) = \sum_{i=0}^{p} \psi_i B^i = \prod_{i=1}^{p} (1 - \phi_i B) \tag{7.5.0.5}$$

for non seasonal cases and

$$\psi_{\P}(B) = \phi_p(B) \; . \; S_s(B) = \sum_{i=0}^{\P} \psi_i B^i = \prod_{i=1}^{\P} (1 - \gamma_i B) \tag{7.5.0.6}$$

for seasonal cases, where $S_s(B)$ is a polynomial in B^s for seasonality, defined in section (6.4.2). In the present study, for the seasonal cases (case 2 and case 4) where the data sets are quarterly (seasonal period n = 4) a damped form (6.4.2.4) is considered so that for full harmonic

$$S_3(B) = (1 + r^2 B^2)(1 + rB) \qquad (7.5.0.7)$$

where r is the damping factor, such that $r = \tilde{b}\,\beta^{1/2}$ for any constant $\tilde{b}$ lying between zero and one, and $0 < r < 1$.

In all cases on-line variance learning (6.8) is used as the actual variances of the time series under study are unknown. For the protection against outliers, modification (6.8.0.5) with confidence factor $\xi = 4$ is considered so that

$$V_t = X_t / N_t$$

$$X_t = \beta_v X_{t-1} + (1 - \underline{u}_t \underline{A}_t) d_t$$

$$d_t = \mathrm{Min}(e_t^2 , 4 \hat{Y}_t)$$

$$N_t = \beta_v N_{t-1} + 1 \qquad (7.5.0.8)$$

In the non-seasonal cases (case 1 & case 3) variance learning is used after 3+p realizations and in seasonal cases after 3+¶ realizations. The priors considered for the learning system are described in the respective case studies.

In all cases the prior for the parameter $\underline{\theta}_t$, considered is

$$\underline{\theta}_0 \sim N(\hat{\underline{\theta}}_0, \underline{C}_0)$$

where $\hat{\underline{\theta}}_0 = \underline{m}_0 = (m \quad b)'$ is set asymptotically as

$$m \cong g /(1 - \lambda) \qquad (7.5.0.9)$$

where g is a rough estimate of the growth rate of the data. The value of b is found such that m+b gives us a value that approximately represent the first few realizations.

$$\underline{C}_0 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$ is used due to precise knowledge of

the positions of $\theta_1 + \theta_2$ and $\theta_1 - \theta_2$ at time $t = 0$. However, this setting is not unique. Other settings such as

$$\underline{C}_0 = \begin{bmatrix} 1.1 & -1 \\ -1 & 1 \end{bmatrix}$$

provide equivalently good forecasts. For results in such cases and more discussion see Akram-Harrison (1983). Similarly, the procedure for setting $\underline{m}_0$ is not unique. Sometime it is convenient to set $\underline{m}_0$ and $\underline{C}_0$, considering the dynamic system in a canonical form, such as

$$\underline{m}_{0,can.} = (\hat{y}_0 \quad g)' \quad \text{and} \quad \underline{C}_{0,can.} = \text{diag}(c, c)$$

where $\hat{y}_0$ is a rough estimate of a value prior to the first observation, g is the growth rate and c is any constant greater than zero; and transform the canonical set up to a set up for a dynamic system in a diagonal form as

$$\underline{m}_0 = \underline{H}\,\underline{m}_{0,can.} \quad \text{and} \quad \underline{C}_0 = \underline{H}\,\underline{C}_{0,can.}\underline{H}'$$

following the transformation and similar models procedure described in section (6.6) of chapter six.

7.5.1 Case 1

An annual data set consisting of 80 observations on the availability of electricity in G.B. (1901-1980) is analysed. First an E.W.R. version of Normal Discount Bayesian Model (7.5.0.1), i.e.

$$Y_t = \underline{f}\,\underline{\theta} + \delta_t \quad ; \quad \delta_t \sim N(0, V_t) \qquad (7.5.1.1)$$

is applied, using the prior

$$\underline{\theta}_0 \sim N\left(\begin{bmatrix} 8 \\ -9 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right)$$

Following the procedure explained earlier the value of λ is found approximately equal to 0.983. The stated prior value of $\underline{m}_0$ is found using (7.5.0.9) with $g \simeq 0.14$. For low frequency the discount factor considered is $\beta = 0.96$. $\underline{f}$ and $\underline{G}$ are

$$\underline{f} = (1 \quad 1) \quad ; \quad \underline{G} = \text{diag}(1, 0.983)$$

The variance of the data is unknown, so on line variance learning (7.5.0.8) is used, replacing $\underline{u}_t$ by $\underline{f}$ and considering the prior $V_0 = 1$, $N_0 = 5$ and discount factor β_v=0.99. The estimated values of the variance V_t are used after the first three observations as no significant contribution from the estimated variances is expected for the first three observations.

One step ahead forecasts and residuals are obtained by transforming the original data to the log form, using the recurrence relations (3.3.2) and following the Forecasting Schemata 2 (7.3). The observations and the one step ahead forecasts are displayed in Fig.4a. The Average String Length of the residuals or one step ahead forecast errors is evaluated, and is found to be approximately equal to 14. This value, obviously quite significantly different from 2, falls in the critical region 2 (C.R.2: region > 2.6). This suggests a value of AR(1) coefficient ϕ around 0.9, on comparison with the theoretical values of A.S.L. (Appendix A). This shows the presence of high frequency or Coloured Noise in the residuals, which is even observable in Fig.4a. This means that the E.W.R. type model is inadequate for the data under study.

Following the Stepwise Identification Procedure (S.I.P.) we disregard the model (7.5.1.1) and proceed to G.E.W.R. type models and apply the Normal Discount Bayesian Model (7.5.0.1)

$$Z_t = \underline{u}_t \, \underline{\theta} + \delta_t \qquad ; \delta_t \sim N(0, V_t). \qquad (7.5.1.2)$$

The AR(1) form of this model is applied to the data by considering $\phi = 0.9$, the same prior (as used for the E.W.R. type model) and by using variance learning (7.5.0.8). Z_t and $\underline{u}_t$ are found by defining $\psi(B) = (1 - 0.88B)$. One step ahead forecasts are obtained by using the recurrence relations (6.4.0.4) and are shown in Fig.4b. The A.S.L. value of the residuals incurred by this model is again evaluated and is found to be approximately equal to 2.61 Applying the test of significance (7.2.3), this empirical value is found to be significantly different from 2 and lies in the critical region 2. Comparing with the theoretical values of A.S.L., the estimated value of $\phi \cong 0.36$. In the light of this information AR(1) form of the model (7.5.1.2) seems to be not fully capable of handling the data under study.

Following S.I.P., the AR(1) type N.D.B.M. is rejected and we proceed further to AR(2) type N.D.B.M. by retaining ϕ as ϕ_1 for AR(2) process and considering $\phi = 0.5$ (A value slightly higher than the value recommended by A.S.L. This value gives slightly better forecasts). Z_t series and $\underline{u}_t$ vectors are derived by defining

$$\psi(B) = (1 - 0.88B)(1 - 0.49B)$$

This AR(2) form of the model (7.5.1.2) is applied to the data using the prior, same as in AR(1) case. This time the model selected yielded forecasts with A.S.L. of the residuals not significantly different from 2. One step ahead forecasts obtained from this model are displayed in Fig.4c.

After making sure that the residuals are free from the coloured noise, we find long term (up to 20 steps ahead) forecasts and the trends, following the Forecasting Schemata 2.

Fig.4d shows one step ahead trend values. This gives us an idea about the level underlying the stochastic phenomena of the availability of electricity in G.B.

The 20 steps ahead forecasts and the trend are shown Fig.4e and Fig.4f respectively. Long term forecasts seem to approach the actual trend values. This means that in the joint modelling scheme the low frequency is well protected from the high frequency in the long run.

One step ahead forecasts obtained by the AR(2) form of the N.D.B. Model are optimum in the Minimum Mean Square Error (M.M.S.E.) sense. The Mean Square Error (M.S.E.) and Mean Average Deviation (M.A.D.) found in this case are less than 0.5% and 4.4% respectively.

Long term (20 steps ahead) forecasts are found from the forecast function (6.4.1.2), with $p = 2$ and the trends are obtained from the forecast function (3.3.2.8), i.e.

$$F_t(k) = \underline{f}\ \underline{G}^k\ \underline{m}_t \qquad \text{for all } k \geq 1 \qquad (7.5.1.3)$$

Availability of Electricity in G.B. (1901-1980)

One Step ahead Full Forecasts

E.W.R. type model

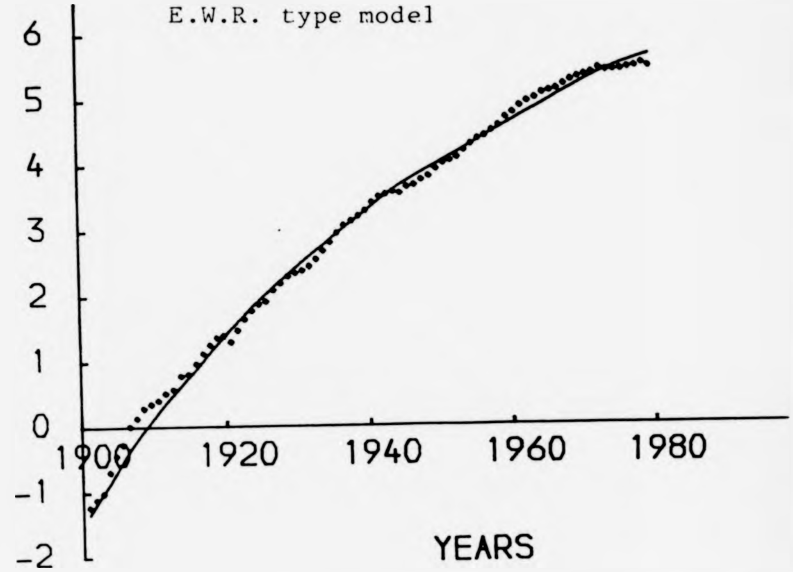


Fig.4a

AR(1) G.E.W.R. type model

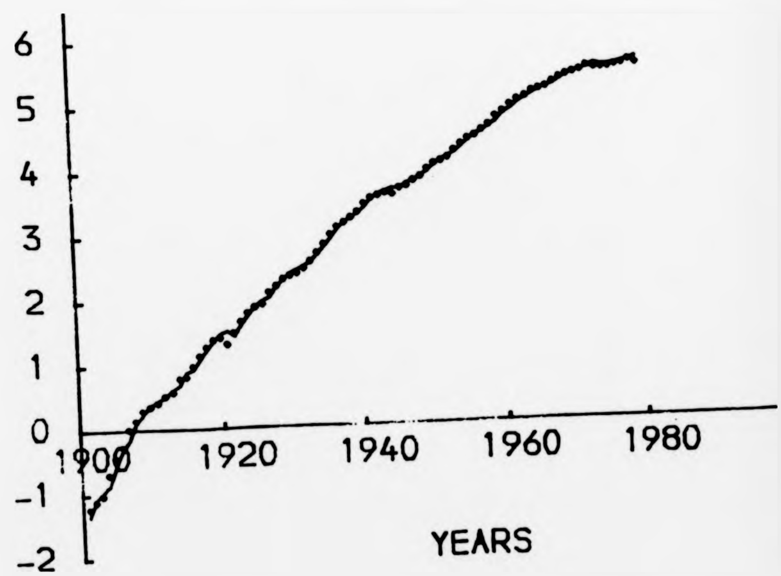


Fig.4b

Dots: Observations

Line: Forecasts (One year a head)

One Step ahead Full Forecasts
AR(2) G.E.W.R. type model

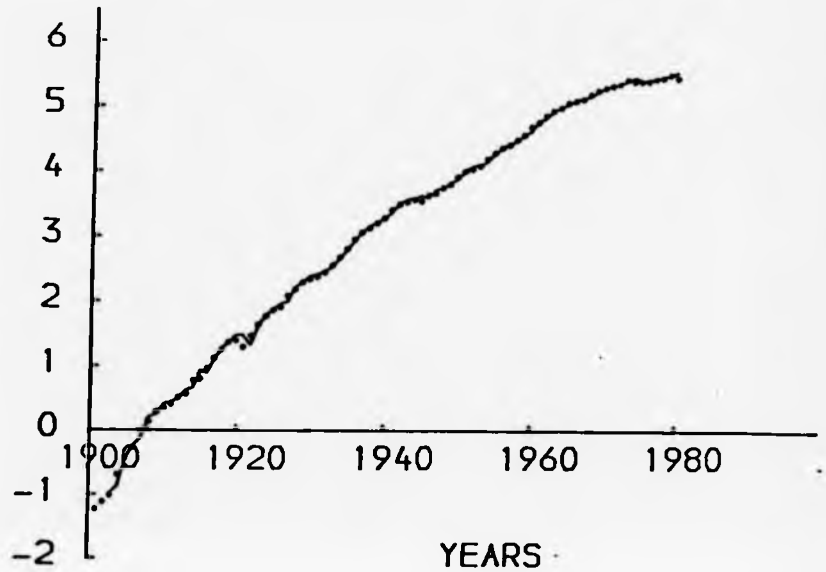


Fig.4c

One Step ahead Trend

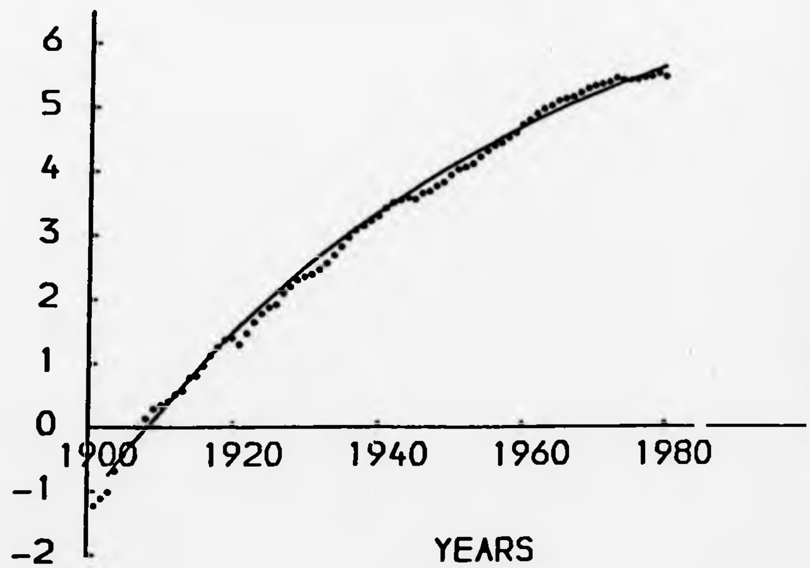


Fig. 4d

Dots: Observations
Line: Forecasts (One year a head)

20-Steps ahead Full Forecasts
AR(2) G.E.W.R. type model

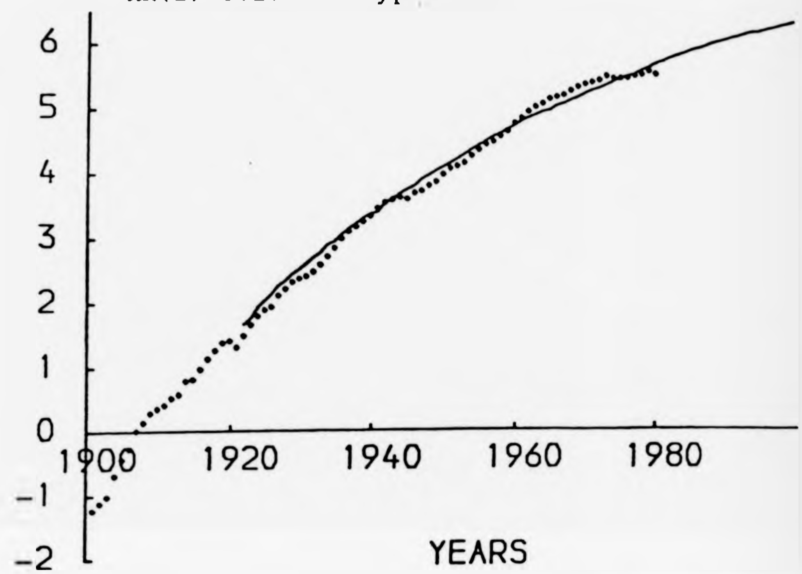


Fig.4e

20-Steps ahead Trend

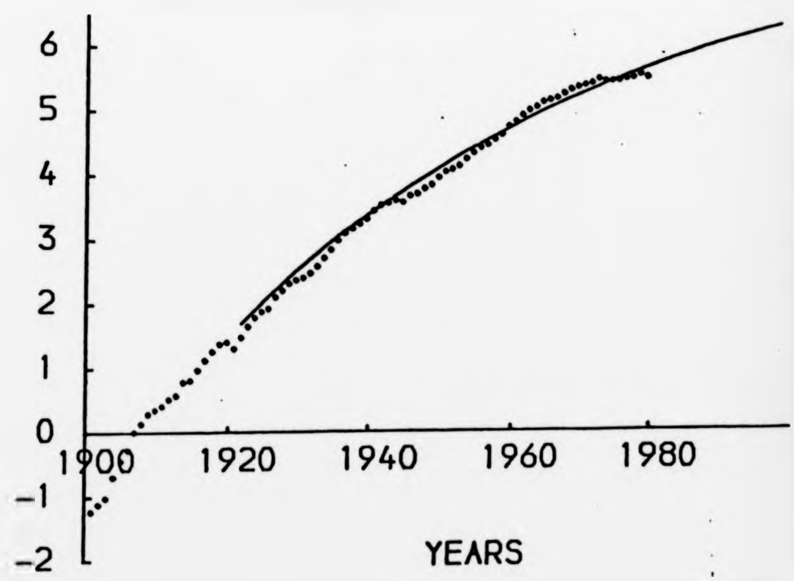


Fig.4f

Dots: Observations
Line: Forecasts (20 years ahead)

7.5.2 Case 2

A quarterly seasonal data series of unadjusted indices of the Chemical Industries of U.K., consisting of 123 observations from 1952 to 1982, is analysed following the Stepwise Identification Procedure (S.I.P.) (7.2.4).

First an E.W.R. version of the Normal Discount Bayesian Model (N.D.B.M.)(7.5.0.1)

$$Z_t = \underline{u}_t \, \underline{\theta} + \delta_t \qquad ; \delta_t \sim N(0, V_t) \tag{7.5.2.1}$$

is applied. Z_t and $\underline{u}_t$ are derived considering (7.5.0.6) with $\psi(B) = 1$,

$$\psi_3(B) = S_3(B) = (1 + r^2B^2)(1 + rB), \tag{7.5.2.2}$$

$\beta = 0.98$, $r = 0.99\ \beta^{1/2}$, $\underline{f} = (1 \ \ 1)$ and $\underline{G} = \text{diag.}(1, \lambda)$

where $\lambda = 0.993$ is evaluated following the procedure explained earlier. The prior distribution of $\underline{\theta}$, at time $t = 0$, used is

$$\underline{\theta}_0 \sim N\left[\begin{pmatrix} 6.3 \\ -3 \end{pmatrix} , \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right]$$

On line variance learning is used with prior

$$\beta_v = 0.99, \ V_0 = 1 \ \text{ and } \ N_0 = 5$$

the same as is used in case 1. The variances estimated are used by the system after first 6 observations.

One step ahead forecasts and residuals are found by using the recurrence relations (6.4.0.4). The Average String Length (A.S.L.) evaluated for the residuals is found to be about 13. This value which is obviously quite significantly different from 2 lies in the critical region 2 (C.R.2: region > 2.45 at $\alpha = 0.05$) and suggests the presence of coloured noise in the residuals with an AR(1) coefficient of around = 0.9. In the light of this information the E.W.R. type version of the model is rejected and we proceed to the G.E.W.R. version of (7.5.2.1) with

$$\psi_4(B) = (1 -0.89B)(1 + r^2B^2)(1 + rB) \qquad (7.5.2.3)$$

Z_t and $\underline{u}_t$ are derived using $\psi_4(B)$. Forecasts yielded by this model are again found unsatisfactory on the suggestion of A.S.L. found from the residuals incurred by this model. The A.S.L. found approximately equal to 5 significantly different from 2 suggested to go ahead to AR(2) version of the N.D.B.M. This new version of the N.D.B.M. with

$$\psi_5(B) = (1 -0.39B)(1 -0.49B)(1 + r^2B^2)(1 + rB) \qquad (7.5.2.4)$$

is applied to the data. This time the model selected performed quite reasonably in general even due to very irregular behaviour after 1972. The A.S.L. value for the span 1952 to 1972 is found to be not significantly different from 2. The time span 1973 to 1982 is not included for the evaluation of A.S.L. evaluated from the residuals incurred by the AR(2) version of the N.D.B.M. due to the very irregular behaviour of the stochastic phenomena. In fact, intervention analysis is required for this part of the data, which is not applied intentionally in order to investigate the over all performance of the model.

The one step ahead forecasts (regular line) plotted against the log transformed data (dots) are displayed in Fig.5a. The one step ahead trend is shown in Fig.5b. Five years ahead (20 steps ahead) forecasts and trend are shown in figures 5c and 5d respectively. The long term forecasts and trend provide a reasonable outlook on the future of Chemical Industries in U.K., though seem to be slightly optimistic (due to non intervention). However, the long term forecasts are in line with the growth rate predicted by many economic forecasters.

The Mean Square Error and Mean Average Deviation found for the one step ahead forecast errors are less than 0.16 % and 2.8 % (for whole data) respectively.

The k-steps ahead forecasts are found from the forecast function (6.4.1.2) by considering $\psi_5(B)$ (7.5.2.4) set up for the model (7.5.2.1) and replacing p by $\zeta = 5$.

The trends (short term and long term) are obtained from the forecast function (7.5.1.3) defined in case 1.

The k-steps ahead forecasts are found from the forecast function (6.4.1.2) by considering $\psi_5(B)$ (7.5.2.4) set up for the model (7.5.2.1) and replacing p by $\zeta = 5$.

The trends (short term and long term) are obtained from the forecast function (7.5.1.3) defined in case 1.

U.K. Chemical Industries Indices (1952-1982)

One Step ahead Full Forecasts
AR(2) G.E.W.R. type model

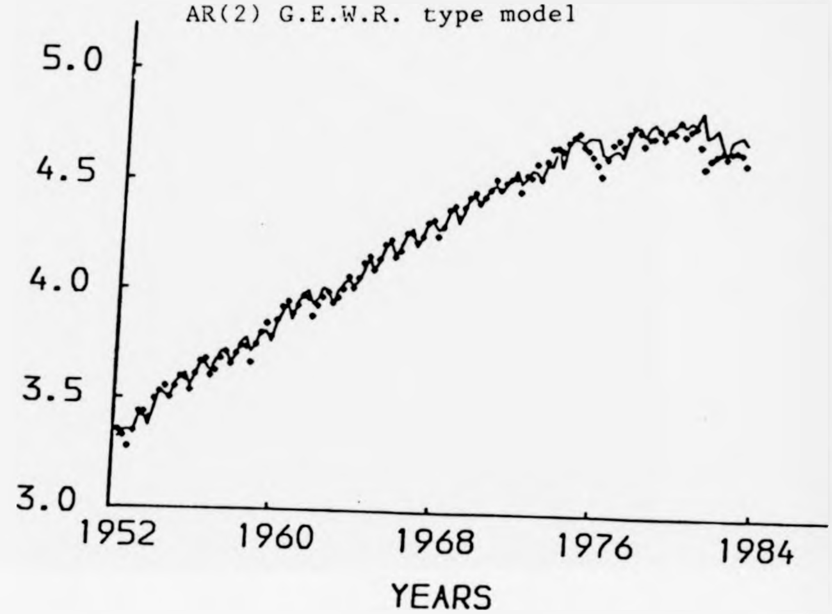


Fig. 5a

Dots: Observations
Line: Forecasts (One quarter ahead)

One Step ahead Trend

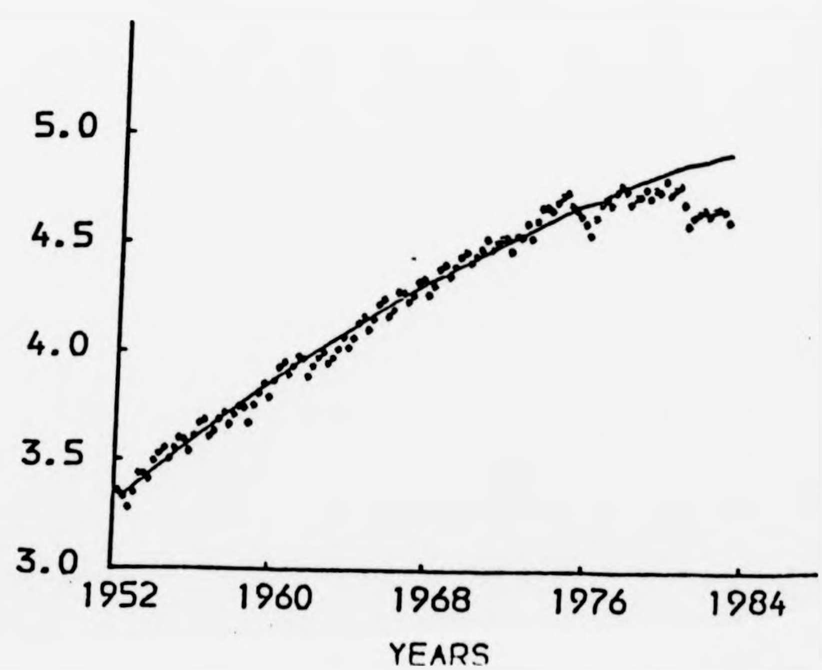


Fig. 5b

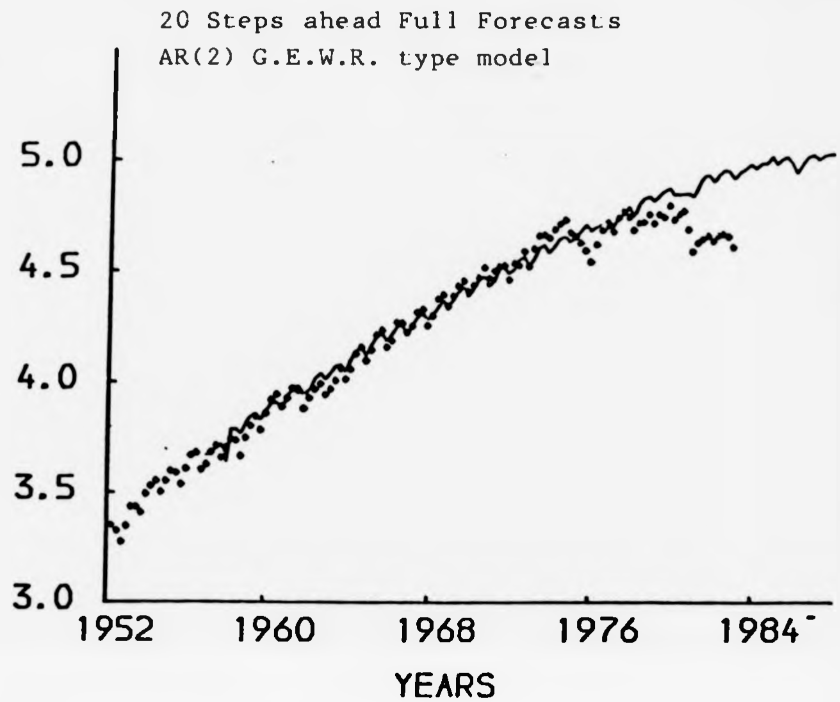


Fig. 5c

Dots: Observations
Line: Forecasts (Five years ahead)

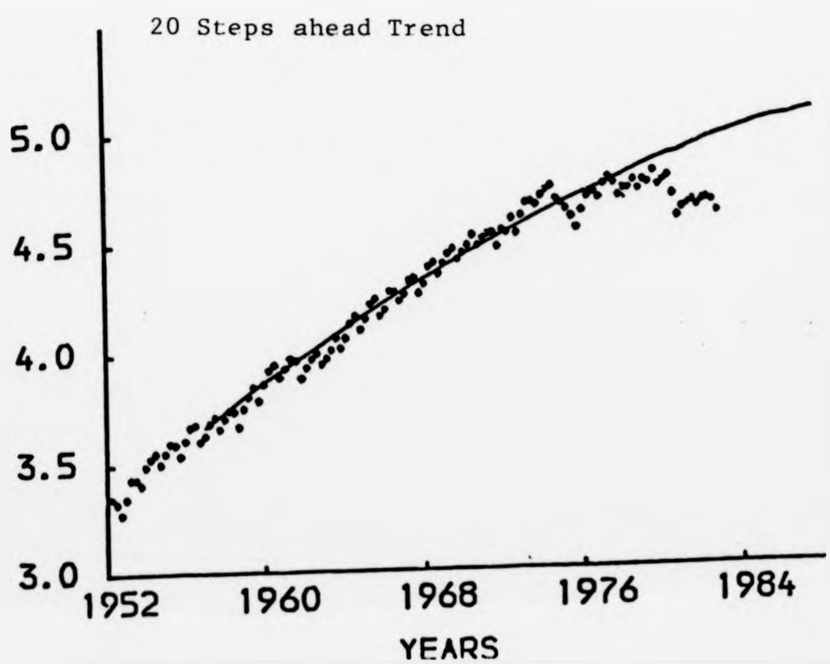


Fig. 5d

7.5.3 Case 3

An annual data set consisting of 82 observations of the population of U.S.A. from 1900 to 1981 is analysed by considering the set up of case 1. Most of the ingredients of case 1 such as

$$\underline{f} = (1 \quad 1) \; , \; \underline{G} = \text{diag.}(\, 1,\ 0.983 \,) \; , \; \beta = 0.96$$

$$\beta_v = 0.99 \; , \; N_0 = 5 \text{ and } V_0 = 1$$

are used here. The prior distribution of $\underline{\theta}_t$ considered is

$$\underline{\theta}_0 \sim N\left[\begin{pmatrix} 5.8 \\ -1.6 \end{pmatrix} \; ; \; \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right]$$

The models (from E.W.R. type to G.E.W.R. type) are identified by the Stepwise Identification Procedure (S.I.P.) and applied to the data as in case 1. On line variance learning is used by the dynamic system after the first three observations.

First an E.W.R. type model D.E.W.R. (3.3.2)

$$Y_t = \underline{f}\,\underline{\theta} + \underline{\delta}_t \qquad ; \; \delta_t \sim N(\, 0,\ V_t \,) \qquad (7.5.3.1)$$

is applied using the above setting. One step ahead forecasts along with residuals are found. A.S.L. ≅ 40 , evaluated from the residuals (obviously significantly different from 2) suggested the value of AR(1) coefficient around 0.9, on comparison with the theoretical values of A.S.L. given in Appendix A. On the bases of this information, it was decided to proceed to a G.E.W.R. type Normal Discount Bayesian Model (7.5.0.1)

$$Z_t = \underline{u}_t\,\underline{\theta} + \delta_t \qquad ; \; \delta_t \sim N(\, 0,\ V_t \,) \qquad (7.5.3.2)$$

where Z_t and $\underline{u}_t$ are derived defi ing

$$\psi_1(B) = (\, 1 - 0.89B \,).$$

This model yielded one step ahead forecasts with residuals having A.S.L. = 3, a value significantly different from 2. This value lies in the critical region 2 (region >2.6 at $\alpha = 0.05$).

Following S.I.P. AR(2) version of the model (7.5.3.2) is considered with

$$\psi_2(B) = (1 - 0.89B)(1 - 0.49B)$$

This model yielded quite good forecasts (Fig.6a) with A.S.L. of residuals not significantly different from 2. The forecasts (one step ahead) are optimum in the Minimum Mean Square Error Sense with Mean Square Error and Mean Average Deviation less than 0.02 % and 0.5 % respectively. After making sure that the residuals are free from the coloured noise, we find the long term (20 steps ahead) forecasts and the trend, which are shown in figures 6c and 6d respectively.

Figure 6b displays one step ahead trend or low frequency. Long term forecasts (Fig. 6c) closely resemble the long term trend which clearly shows that in the long run the trend or low frequency is well protected from the high frequency in our scheme of jointly modelling the low and high frequency componenets, and the model identified by the S.I.P. is equivalently good for short and long term forecasts. along with the trends.

The population of U.S.A. seems to be swinging around the trend. The demographic cycle formed is perhaps due to two world wars. After 2000 A.D., the demographic cycle seems to be repeating its cycle and crossing the long term trend, perhaps a signal for a third world war !.

The k-steps ahead forecasts and trends are found from the forecast functions as in case 1.

Population of U.S.A. (1900 - 1981)

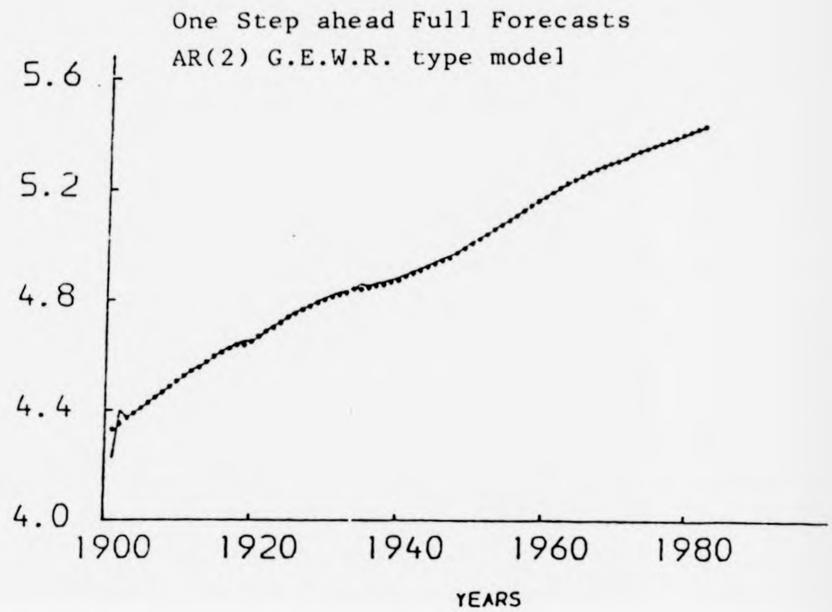


Fig. 6a

Dots: Observations

Line: Forecasts (One year ahead)

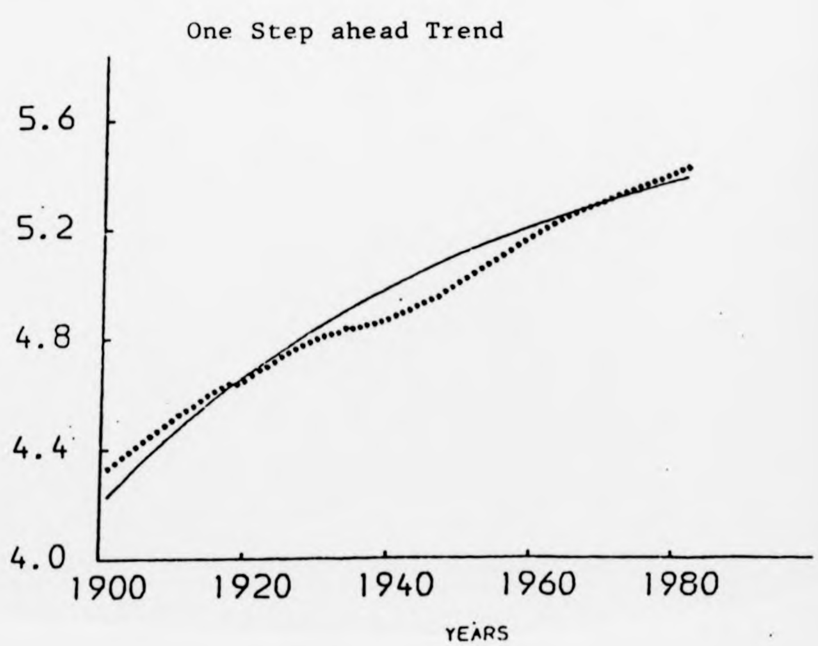


Fig. 6b

20 Steps ahead Full Forecasts
AR(2) G.E.W.R. type model

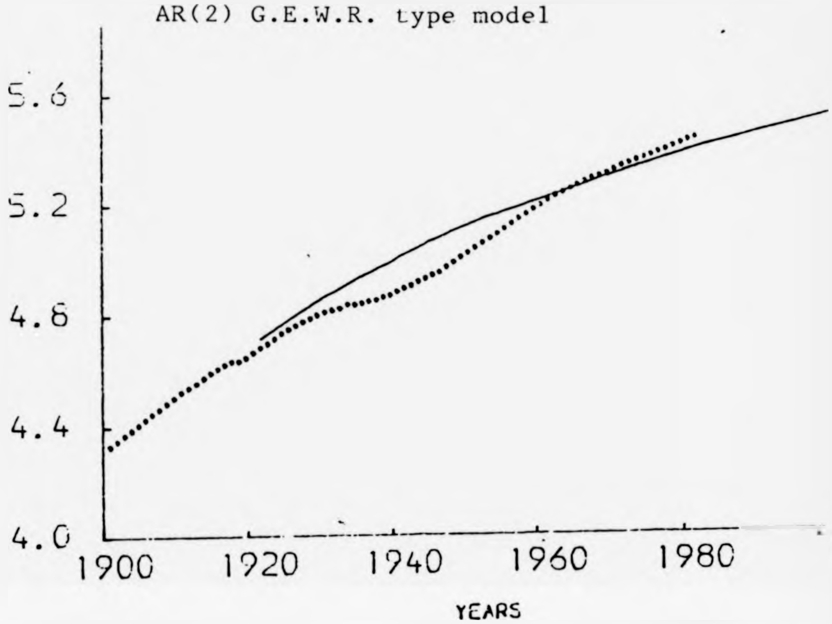


Fig. 6c

Dots: Observations
Line: Forecasts (Twenty years ahead)

20 Steps ahead Trend

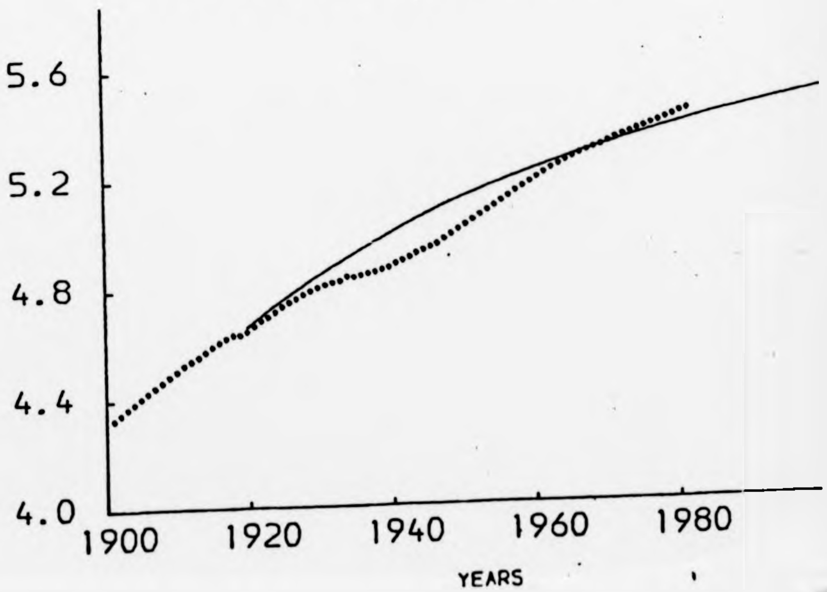


Fig. 6d

7.5.4 Case 4

A quarterly seasonal data set concerned with the disposable personal income in Austria, consisting of 104 realizations of a time series from 1954 to 1979 is analysed by considering a set up for the models similar to case 2.

First, as usual, an E.W.R. type model, i.e. an E.W.R. version of Normal Discount Bayesian Model (N.D.B.M.)(7.5.0.1)

$$Z_t = \underline{u}_t \, \underline{\theta} + \delta_t \quad ; \delta_t \sim N(0, V_t) \qquad (7.5.4.1)$$

is applied with setting

$$\underline{f} = (1 \quad 1) \, , \, \underline{G} = \text{diag.}(1, 0.994), \beta = 0.98,$$

$$r = 0.99 \, \beta^{1/2}$$

and prior distribution of $\underline{\theta}$ as

$$\underline{\theta} \sim N \left[\begin{pmatrix} 6 \\ -3 \end{pmatrix} ; \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right]$$

Z_t and $\underline{u}_t$ are derived by considering $\psi(B) = S_3(B)$ as defined in case 2.

On line variance learning is used with the usual prior $N_0 = 5$ & $V_0 = 1$ and the discount factor $\beta_v = 0.99$.

One step ahead forecasts along with the residuals or one step ahead forecast errors are obtained using the recurrence relations (6.4.0.4). An A.S.L. $\simeq$ 17 is found from the residuals, showing the inadequacy of the E.W.R.type model for the data being analysed. The A.S.L. value which is quite significantly different from 2 suggested the value of AR(1) coefficient ϕ around 0.9. Following the S.I.P. we decided to proceed to AR(1) version of the above model, similar to the model considered in case 2. In order to obtain Z_t and $\underline{u}_t$, $\psi(B)$ defined in case 2 is considered with $\phi = 0.9$.

A N.D.B.M. with an AR(1) component applied to the data yielded one step ahead forecasts with residuals having an A.S.L. equal to 4.3, a value not significantly different from 2 at the 5% level of significance and as usual lying in the critical region 2 (C.R.2: region > 2.51 at $\alpha = 0.05$). In the light of this information and following S.I.P. , an AR(2) version of the model (7.5.4.1) is selected with $\psi_5(B)$ as defined in case 2.

The AR(2) version of the model (7.5.4.1) yielded optimum forecasts in the Minimum Mean Square Error sense and A.S.L. not significantly different from 2. One step ahead forecasts found are with residuals forming a White Noise sequence, and having Mean Square Error and Mean Average Deviation less than 0.06 % and 1.6 % respectively.

One step ahead forecasts along with the observations are displayed in Fig.7a. One step ahead trend is shown in Fig. 7b.

Long term forecasts (5 years or 20 steps ahead) and long term trend is displayed in Figures 7c and 7d respectively.

k steps ahead forecasts and trends (both short term and long term) are obtained from the forecast functions considered in case 2.

Austrian Disposable Personal Income (1954-1979)

One Step ahead Full Forecasts
AR(2) G.E.W.R. type model

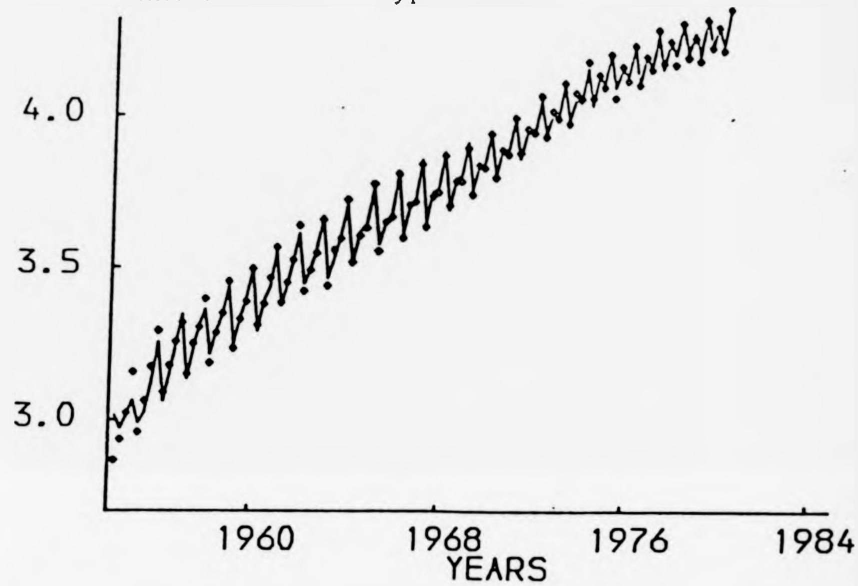


Fig. 7a

One Step ahead Trend

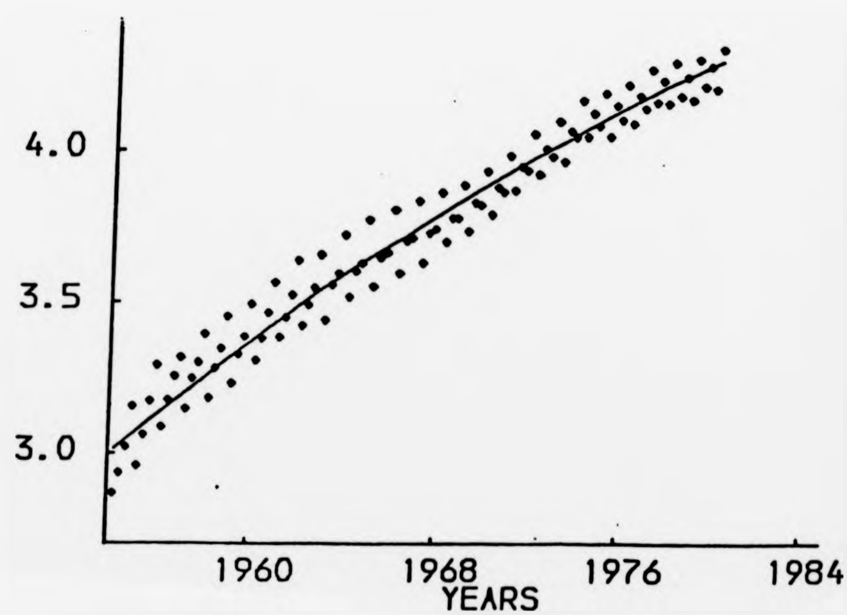


Fig. 7b

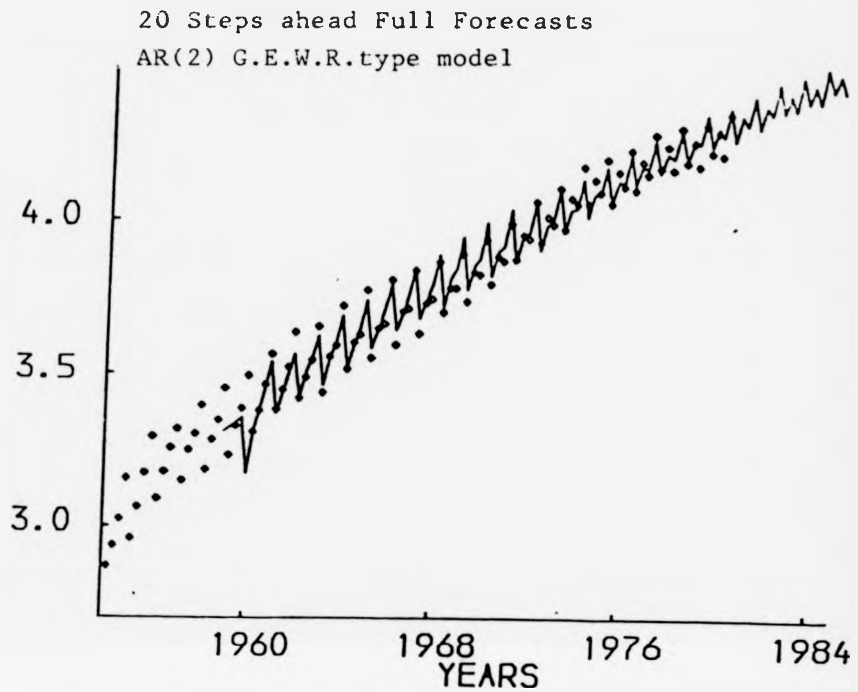

20 Steps ahead Full Forecasts
AR(2) G.E.W.R.type model
4.0
3.5
3.0
1960
1968
1976
1984
YEARS

Fig. 7c

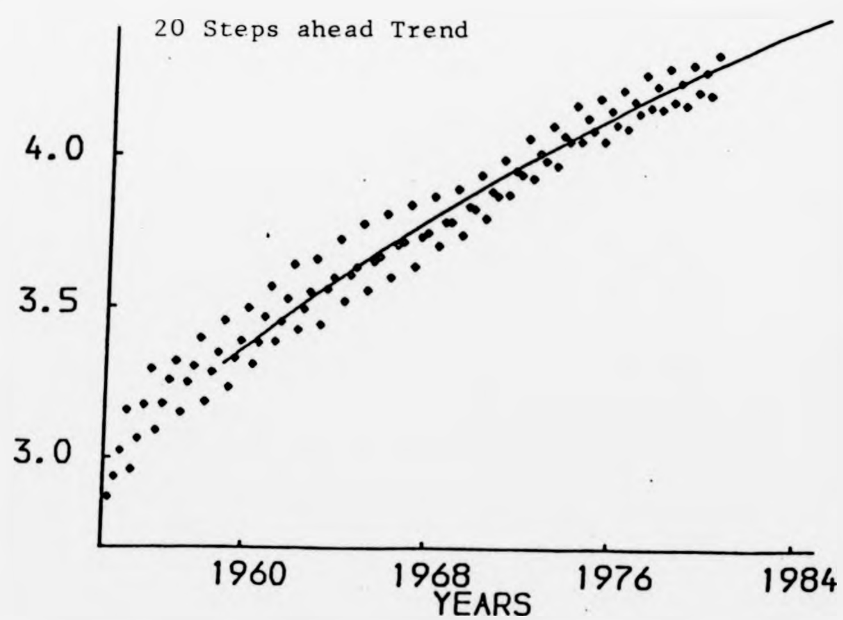

20 Steps ahead Trend
4.0
3.5
3.0
1960
1968
1976
1984
YEARS

Fig. 7d

7.6 Comment

The four cases of real life data sets studied clearly show the capability of G.E.W.R. methodology to handle stochastic phenomena driven by coloured noise processes, quite reasonably. In all cases the E.W.R. type models, initially applied, are found inadequate to filter the coloured noise; whereas, the G.E.W.R. type models ultimately applied, gave us forecasts with one step ahead residuals or forecast errors forming a white noise process, i.e. having A.S.Ls. not significantly different from 2.

The Stepwise Identification Procedure (S.I.P.), in all cases, properly identified the models. It is, however, a matter of coincidence that in all cases the ultimate models identified are AR(2) G.E.W.R. type models with AR coefficients quite close to each other, even due to the fact that the data sets not only belong to different environments but also belong to different walks of life. In all cases, for the high frequency component or coloured noise the AR coefficients used are $\phi_1 = 0.9$ and $\phi_2 = 0.5$. The forecasts obtained by using these coefficients in all cases are optimum in both the senses of Minimum Mean Square Error (M.M.S.E.) and the whiteness of the residuals (A.S.Ls, not significantly different from 2). The G.E.W.R. models with these coefficients, which proved to be quite suitable for all four data sets, may be suitable in many real life situations, but it is difficult to say or establish a rule that G.E.W.R. type models with such settings will in general be true for time series driven by coloured noise processes. In this direction, still lot of work is required to establish such a thing. In higher order coloured noise cases, it is hoped that the S.I.P. will help us to identify coloured noise present in time series and assist in selecting correct forecasting model.

The joint modelling scheme, introduced within the concept of G.E.W.R., behaved well as is clear from the short and long term forecasts in all cases. In the long run the low frequency is well protected from the high frequency.

In all cases the on-line variance learning system introduced improved the performance of the models quite significantly.

8
CONCLUSION AND SUGGESTIONS FOR FURTHER RESEARCH

8.1 Conclusion and Remarks

A general development of the theory of Generalised Exponentially Weighted Regression (G.E.W.R.) is presented along with practical implications.

The infrastructure required for the development of the theory is reviewed, discussed and developed further before concentrating on the main nucleus of the research. The work presented is confined to the development of the theory for the analysis and forecasting of discrete time oriented stochastic processes driven by ARMA type Coloured Noise processes.

Most of the work presented is relevant to the two joint papers, Harrison-Akram (1983) and Akram-Harrison (1983), except for some new results developed during the gap between the presentation of the second paper and the writing of the thesis.

A very detailed and technically concise theory of G.E.W.R., as remarked by Mckenzie (1983), has successfully been applied not only to the simulated data sets but also to the real life economic, social and industrial time series driven by the Coloured Noise processes. In order to demonstrate the capability and validity of the theory, G.E.W.R. type Dynamic Linear Models are applied to simulated data sets after going through the Stepwise Identification Procedure (S.I.P.)(7.2.4) and G.E.W.R. type Normal Discount Bayesian Models for the real life time series. In the real life cases the actual variances were not known, so an on-line Bayesian learning procedure was used to estimate the variances. This learning system improved the performance of the models quite significantly.

In all case studies (simulation and real life) a joint modelling scheme, i.e. jointly modelling the low, medium and high frequencies within the same framework, is adopted. The models selected following the S.I.P. yielded not only optimum one step ahead forecasts in the minimum mean square error sense and the Whiteness of the residuals through the Average String Lengths (A.S.L.); but also gave us quite reasonable long term forecasts and trends. In all cases it is quite evident that in the long run the low frequency is well protected from the high frequency.

The theorems along with the corollaries are presented in a fairly simple and straightforward manner. The theorem (T3) not only generalises the result of Mckenzie (1976) but also provide a simple , precise and general solution.

The recurrence relations (6.6) developed for the transformation of a dynamic system in a canonical form to a dynamic system in a diagonal form (and vice-versa) have their own significance,especially when we search for similar models,as the recurrence relations developed for the transformation matrices are independent of matrix inversions.

It is well known that a linear filter is optimum if the system is asymptotically stable and the system is asymptotically stable if all its weighting functions decay to zero as $t\to\infty$. This highlights the significance of an exponential weighting system in linear filtration. The importance increases further when the whole set up is considered within a Bayesian framework and we introduce the coloured noise structure to deal with general situations. All these are basic ingredients of the G.E.W.R. theory developed.

Various parsimonious Bayesian Dynamic Linear Models and Normal Discount Bayesian Models, introduced for the low and high frequency componenets of time series with or without seasonality and cyclical fluctuations are fairly straightforward and simple to use. The models introduced provide robust forecasts.

A procedure to construct G.E.W.R. type Dynamic Linear Models of canonical form (6.5) is described, which helps us to formulate the models in terms of original series Y_t in place of the derived series Z_t. For complex stochastic systems, a method to construct State Spece models (6.7) is given.

The introduction of this new forecasting methodology has not only opened doors to many new research projects in this area but has also provided stimulus to the general theory of the dicounted weighted models, such as the work of Ameen-Harrison (1983).

8.2 Suggestions For Further Research

There is a tremendous scope for further research in this area, which may be highlighted as:

i) Variable Discount Factor

In the present development of the theory of G.E.W.R., the discount factor β considered is time invariant and only one discount factor is considered for the low and high frequency components . In order to advance this work , the present theory of G.E.W.R. may be developed further to accommodate different discount factors for low and high frequency components, either time invariant or time variant

ii) ARMA Coefficients

Coefficients for the ARMA type Coloured Noise process are assumed constant and identifiable by the procedures such as Autocorrelation Function, Partial Autocorrelation Function and Stepwise Identification Procedure (S.I.P.) based on Average String Lengths, etc. An on-line Bayesian learning procedure for the parameters could be developed on the principles of the variance learning procedure introduced.

iii) Limiting Values of Updating Vectors $\underline{A}_t$

Dobbie's theorem (1963) for the limiting values of the updating vectors is restricted to unequal roots of the transition matrices $\underline{G}$ of diagonal form. A general result can be developed for the dynamic systems in a canonical or any other form with transition $\underline{G}$ having some , or all, roots equal.

iv) $\underline{W}$ Matrices

The $\underline{W}$ matrices considered in the case of Dynamic Linear Models are time invariant in nature. A dynamic time variant system may be developed to update $\underline{W}_t$ matrices through some recurrence relations, with the arrival of new information from the environment.

v) Stepwise Identification Procedure(S.I.P.)

The Stepwise Identification Procedure based on Average String Lengths of AR(1) process can be developed further for general ARMA type processes by developing the probability distribution functions for the signs of the ARMA type Coloured Noise processes.

APPENDIX A1

Average String Lengths of an AR(1) Coloured Noise

	ASL		ASL
-0.999	1.01	-0.50	1.50
-0.99	1.05	-0.49	1.51
-0.98	1.07	-0.48	1.52
-0.97	1.08	-0.47	1.52
-0.96	1.10	-0.46	1.53
-0.95	1.11	-0.45	1.54
-0.94	1.12	-0.44	1.55
-0.93	1.14	-0.43	1.56
-0.92	1.15	-0.42	1.57
-0.91	1.16	-0.41	1.58
-0.90	1.17	-0.40	1.58
-0.89	1.18	-0.39	1.59
-0.88	1.19	-0.38	1.60
-0.87	1.20	-0.37	1.61
-0.86	1.21	-0.36	1.62
-0.85	1.21	-0.35	1.63
-0.84	1.22	-0.34	1.64
-0.83	1.23	-0.33	1.65
-0.82	1.24	-0.32	1.66
-0.81	1.25	-0.31	1.67
-0.80	1.26	-0.30	1.68
-0.79	1.27	-0.29	1.68
-0.78	1.27	-0.28	1.69
-0.77	1.28	-0.27	1.70
-0.76	1.29	-0.26	1.71
-0.75	1.30	-0.25	1.72
-0.74	1.31	-0.24	1.73
-0.73	1.32	-0.23	1.74
-0.72	1.32	-0.22	1.75
-0.71	1.33	-0.21	1.76
-0.70	1.34	-0.20	1.77
-0.69	1.35	-0.19	1.78
-0.68	1.35	-0.18	1.79
-0.67	1.36	-0.17	1.80
-0.66	1.37	-0.16	1.81
-0.65	1.38	-0.15	1.82
-0.64	1.39	-0.14	1.84
-0.63	1.40	-0.13	1.85
-0.62	1.40	-0.12	1.86
-0.61	1.41	-0.11	1.87
-0.60	1.42	-0.10	1.88
-0.59	1.43	-0.09	1.89
-0.58	1.44	-0.08	1.90
-0.57	1.44	-0.07	1.91
-0.56	1.45	-0.06	1.93
-0.55	1.46	-0.05	1.94
-0.54	1.47	-0.04	1.95
-0.53	1.48	-0.03	1.96
-0.52	1.48	-0.02	1.97
-0.51	1.49	-0.01	1.99

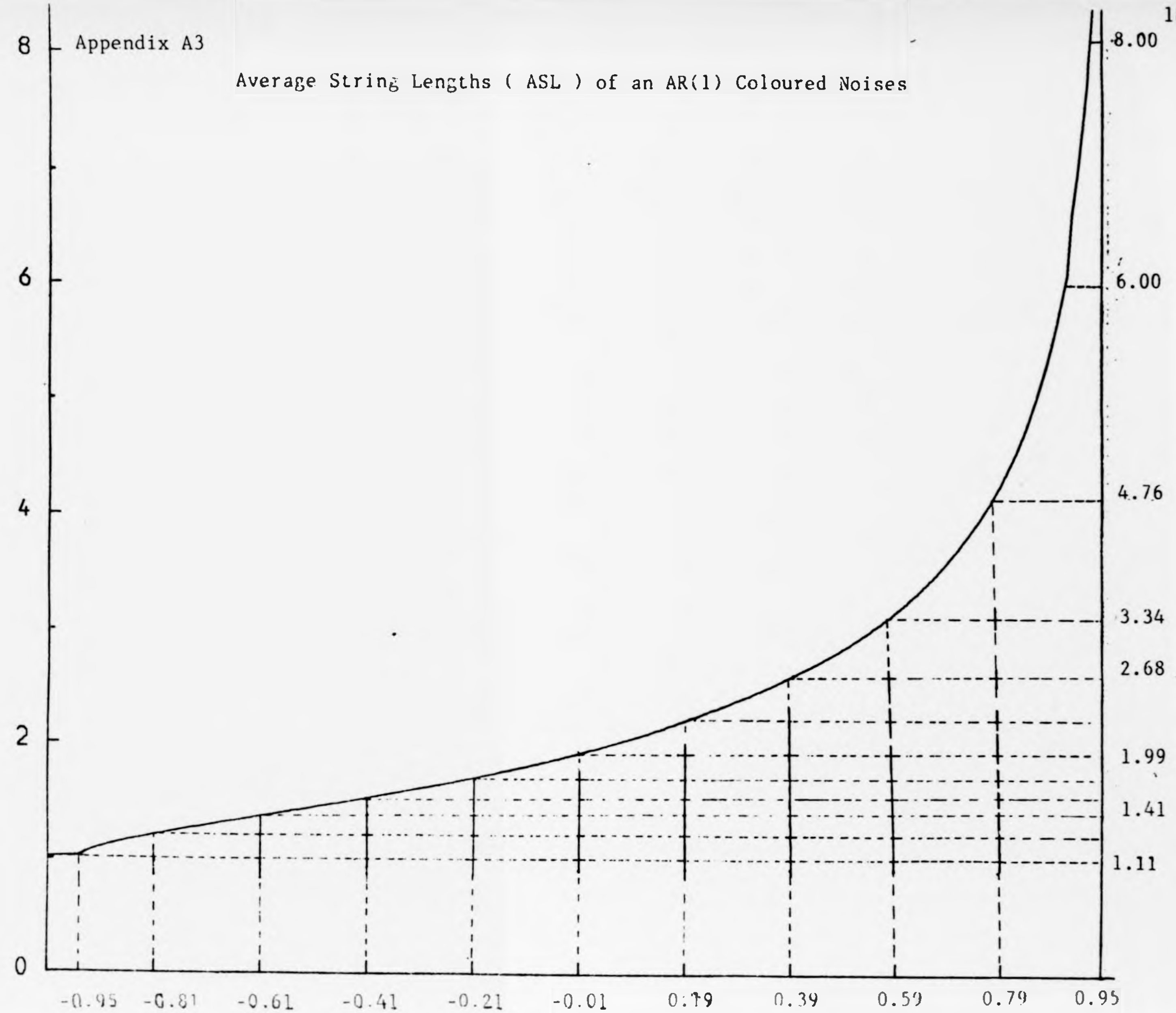
APPENDIX A2

Average String Lengths of an AR(1) Coloured Noise

	ASL		ASL
0.00	2.00	0.50	3.00
0.01	2.01	0.51	3.03
0.02	2.03	0.52	3.07
0.03	2.04	0.53	3.10
0.04	2.05	0.54	3.14
0.05	2.07	0.55	3.18
0.06	2.08	0.56	3.22
0.07	2.09	0.57	3.26
0.08	2.11	0.58	3.30
0.09	2.12	0.59	3.34
0.10	2.14	0.60	3.39
0.11	2.15	0.61	3.43
0.12	2.17	0.62	3.48
0.12	2.18	0.63	3.53
0.14	2.20	0.64	3.59
0.15	2.21	0.65	3.64
0.16	2.23	0.66	3.70
0.17	2.24	0.67	3.76
0.18	2.26	0.68	3.82
0.19	2.28	0.69	3.88
0.20	2.29	0.70	3.95
0.21	2.31	0.71	4.02
0.22	2.33	0.72	4.10
0.23	2.35	0.73	4.18
0.24	2.36	0.74	4.26
0.25	2.38	0.75	4.35
0.26	2.40	0.76	4.44
0.27	2.42	0.77	4.54
0.28	2.44	0.78	4.64
0.29	2.46	0.79	4.76
0.30	2.48	0.80	4.88
0.31	2.50	0.81	5.01
0.32	2.52	0.82	5.16
0.33	2.54	0.83	5.31
0.34	2.57	0.84	5.48
0.35	2.59	0.85	5.66
0.36	2.61	0.86	5.87
0.37	2.64	0.87	6.09
0.38	2.66	0.88	6.40
0.39	2.68	0.89	6.64
0.40	2.71	0.90	6.97
0.41	2.74	0.91	7.35
0.42	2.76	0.92	7.80
0.43	2.79	0.93	8.35
0.44	2.82	0.94	9.02
0.45	2.85	0.95	9.89
0.46	2.87	0.96	11.10
0.47	2.90	0.97	12.80
0.48	2.94	0.98	15.68
0.49	2.97	0.99	22.20
		0.999	70.24

Appendix A3

Average String Lengths (ASL) of an AR(1) Coloured Noises



APPENDIX B1

Linear Growth + AR(1) Coloured Noise with $\phi = 0.3$ from a Random Noise of Mean Zero & Variance 10,000

		6100	154.931	12400	204.526	18700	404.267
		6200	51.7772	12500	308.148	18800	382.190
		6300	238.307	12600	334.622	18900	337.309
100	155.165	6400	168.013	12700	358.893	19000	249.597
200	249.198	6500	192.082	12800	298.696	19100	257.330
300	308.022	6600	39.3907	12900	218.024	19200	264.832
400	165.217	6700	17.7830	13000	54.4010	19300	94.0884
500	278.928	6800	29.1760	13100	234.045	19400	207.500
600	118.093	6900	62.7702	13200	197.781	19500	412.169
700	35.9902	7000	224.975	13300	32.1149	19600	367.936
800	113.001	7100	124.650	13400	364.219	19700	310.365
900	133.933	7200	231.571	13500	356.807	19800	348.422
1000	-26.073	7300	275.416	13600	268.573	19900	339.936
1100	78.1723	7400	71.9746	13700	155.464	20000	264.518
1200	260.354	7500	52.4894	13800	148.225	20100	64.6013
1300	438.395	7600	193.454	13900	16.7987	20200	114.945
1400	277.858	7700	285.461	14000	161.369	20300	122.383
1500	110.075	7800	231.221	14100	252.369	20400	284.265
1600	103.535	7900	329.506	14200	326.666	20500	404.482
1700	259.181	8000	346.653	14300	101.208	20600	105.742
1800	79.6127	8100	-2.3819	14400	0.10122	20700	204.455
1900	154.011	8200	140.249	14500	238.696	20800	128.669
2000	169.442	8300	146.011	14600	59.3265	20900	189.783
2100	150.742	8400	135.256	14700	270.472	21000	181.066
2200	19.7864	8500	225.564	14800	203.760	21100	391.943
2300	205.146	8600	304.908	14900	183.974	21200	529.983
2400	217.902	8700	330.476	15000	189.181	21300	114.673
2500	126.271	8800	290.506	15100	137.301	21400	305.888
2600	351.020	8900	340.822	15200	80.2607	21500	439.690
2700	240.681	9000	222.768	15300	136.574	21600	432.061
2800	212.202	9100	184.260	15400	231.313	21700	295.452
2900	211.171	9200	218.186	15500	251.847	21800	291.866
3000	191.260	9300	314.656	15600	283.587	21900	97.6409
3100	193.420	9400	242.480	15700	194.456	22000	204.674
3200	314.324	9500	312.067	15800	198.187	22100	219.892
3300	232.750	9600	223.494	15900	296.990	22200	351.555
3400	78.0663	9700	318.602	16000	96.5613	22300	543.050
3500	131.146	9800	362.758	16100	209.742	22400	500.983
3600	306.273	9900	321.580	16200	150.310	22500	357.208
3700	72.3117	10000	396.231	16300	196.097	22600	252.521
3800	5.46019	10100	250.409	16400	310.479	22700	354.591
3900	127.143	10200	99.0332	16500	351.228	22800	210.316
4000	-10.417	10300	292.161	16600	375.138	22900	388.613
4100	155.490	10400	146.519	16700	287.795	23000	509.379
4200	191.220	10500	141.652	16800	339.134	23100	160.388
4300	180.761	10600	133.148	16900	242.298	23200	582.781
4400	7.16784	10700	339.681	17000	78.9327	23300	480.715
4500	102.515	10800	335.887	17100	99.1465	23400	373.466
4600	6.22368	10900	365.868	17200	362.096	23500	380.745
4700	97.8625	11000	177.261	17300	259.011	23600	321.122
4800	48.9534	11100	156.357	17400	248.514	23700	250.781
4900	88.7878	11200	50.2049	17500	453.075	23800	420.048
5000	278.790	11300	248.913	17600	247.389	23900	452.078
5100	244.459	11400	251.410	17700	293.390	24000	427.325
5200	130.390	11500	160.825	17800	277.865	24100	317.706
5300	248.549	11600	209.489	17900	323.806	24200	328.698
5400	15.2498	11700	167.024	18000	283.693	24300	306.367
5500	479.998	11800	73.2335	18100	280.582	24400	538.039
5600	372.229	11900	18.5252	18200	397.887	24500	404.407
5700	135.574	12000	187.992	18300	509.211	24600	338.385
5800	105.389	12100	225.143	18400	38.9810	24700	267.645
5900	420.745	12200	82.5149	18500	255.624	24800	343.905
6000	256.274	12300	272.297	18600	317.956	24900	395.736

APPENDIX B2

Linear Growth + AR(1) Coloured Noise with $\phi=0.3$ from a Random Noise of Mean Zero & Variance 10.000

25000	465.962	31300	365.482	37600	360.366	43900	444.090
25100	564.812	31400	451.045	37700	185.758	44000	520.963
25200	396.160	31500	575.845	37800	451.437	44100	446.209
25300	419.194	31600	385.609	37900	602.434	44200	450.580
25400	568.768	31700	506.097	38000	604.304	44300	346.792
25500	450.211	31800	331.067	38100	365.152	44400	347.743
25600	419.842	31900	281.490	38200	563.212	44500	447.668
25700	582.499	32000	384.944	38300	548.401	44600	619.030
25800	387.716	32100	327.783	38400	506.484	44700	603.966
25900	481.376	32200	469.726	38500	627.302	44800	741.128
26000	421.399	32300	431.600	38600	496.079	44900	677.316
26100	540.018	32400	431.402	38700	302.737	45000	643.925
26200	459.617	32500	488.930	38800	474.518	45100	513.315
26300	572.272	32600	553.241	38900	571.129	45200	436.759
26400	433.620	32700	485.109	39000	441.547	45300	456.720
26500	342.392	32800	558.848	39100	410.111	45400	630.927
26600	310.491	32900	627.455	39200	423.298	45500	701.692
26700	482.167	33000	544.975	39300	460.664	45600	676.162
26800	308.726	33100	467.536	39400	589.682	45700	628.409
26900	363.411	33200	469.457	39500	561.230	45800	671.597
27000	273.046	33300	606.366	39600	405.317	45900	537.256
27100	270.507	33400	411.449	39700	452.576	46000	512.817
27200	260.626	33500	328.603	39800	667.255	46100	577.428
27300	389.593	33600	240.714	39900	569.953	46200	471.818
27400	388.156	33700	361.649	40000	567.497	46300	606.547
27500	453.300	33800	374.761	40100	630.598	46400	646.820
27600	531.680	33900	301.732	40200	504.502	46500	546.289
27700	393.129	34000	218.312	40300	470.799	46600	457.353
27800	515.438	34100	541.413	40400	514.933	46700	592.723
27900	538.844	34200	560.586	40500	516.325	46800	510.702
28000	292.905	34300	452.650	40600	365.650	46900	611.842
28100	321.320	34400	330.615	40700	450.934	47000	513.972
28200	414.741	34500	307.819	40800	607.092	47100	383.971
28300	374.653	34600	360.232	40900	401.947	47200	436.150
28400	344.999	34700	442.178	41000	464.678	47300	384.164
28500	365.711	34800	340.225	41100	469.892	47400	550.714
28600	523.878	34900	206.873	41200	323.532	47500	475.968
28700	157.173	35000	243.332	41300	411.510	47600	500.711
28800	94.6988	35100	354.641	41400	582.567	47700	669.139
28900	252.734	35200	335.754	41500	509.317	47800	502.493
29000	231.571	35300	449.455	41600	442.228	47900	549.479
29100	132.372	35400	553.585	41700	418.844	48000	628.809
29200	340.775	35500	254.602	41800	364.346	48100	506.732
29300	250.432	35600	429.621	41900	510.462	48200	696.892
29400	147.508	35700	480.554	42000	485.689	48300	562.973
29500	333.086	35800	517.222	42100	509.056	48400	523.963
29600	471.072	35900	454.955	42200	578.755	48500	675.261
29700	519.724	36000	383.302	42300	491.220	48600	672.013
29800	549.117	36100	453.257	42400	585.758	48700	514.248
29900	487.067	36200	461.719	42500	468.450	48800	489.603
30000	330.810	36300	421.271	42600	479.658	48900	483.679
30100	324.759	36400	530.839	42700	579.614	49000	662.741
30200	45[illegible].976	36500	348.068	42800	637.358	49100	602.297
30300	394.305	36600	356.540	42900	433.466	49200	614.193
30400	365.497	36700	561.617	43000	405.207	49300	503.501
30500	413.169	36800	345.290	43100	747.646	49400	529.093
30600	263.477	36900	247.734	43200	654.502	49500	477.759
30700	235.057	37000	434.738	43300	666.468	49600	482.055
30800	465.789	37100	514.011	43400	519.161	49700	741.744
30900	346.131	37200	568.842	43500	557.687	49800	517.081
31000	570.739	37300	555.005	43600	514.270	49900	530.569
31100	610.882	37400	490.130	43700	482.752	50000	532.943
31200	354.149	37500	442.377	43800	513.149		

APPENDIX_C1

AR(1) Coloured Noise generated with ϕ =-0.9 from a Random Noise of Mean zero & Variance 10,000

		6100	470.303	12400	[illegible]	18700	[illegible]
		6200	-480.66	12500	[illegible]	18800	[illegible]
		6300	526.153	12600	[illegible]	18900	342.454
100	32.332	6400	-510.03	12700	[illegible]	19000	[illegible]
200	37.684	6500	495.930	12800	[illegible]	19100	309.301
300	39.042	6600	-531.81	12900	[illegible]	19200	[illegible]
400	-59.63	6700	[illegible]	13000	[illegible]	19300	[illegible]
500	157.49	6800	-444.85	13100	[illegible]	19400	[illegible]
600	-200.2	6900	382.160	13200	[illegible]	19500	[illegible]
700	159.65	7000	-279.87	13300	198.5661	19600	-165.25
800	-113.2	7100	197.265	13400	-25.7472	19700	145.191
900	[illegible]	7200	-108.21	13500	24.10212	19800	-102.75
1000	[illegible]	7300	128.823	13600	-17.0138	19900	96.9709
1100	206.69	7400	-200.02	13700	-33.6034	20000	-118.07
1200	-103.4	7500	161.304	13800	15.63037	20100	-18.306
1300	233.54	7600	-100.96	13900	-119.149	20200	-10.089
1400	-217.1	7700	141.828	14000	143.0834	20300	-38.709
1500	170.43	7800	-127.81	14100	-116.106	20400	94.0993
1600	-146.5	7900	201.429	14200	152.3592	20500	-40.374
1700	220.74	8000	-133.26	14300	-250.133	20600	-120.07
1800	-271.1	8100	-29.279	14400	149.8543	20700	134.971
1900	306.41	8200	94.3291	14500	-64.6636	20800	-224.57
2000	-269.2	8300	-118.10	14600	-74.1516	20900	194.853
2100	253.99	8400	101.678	14700	164.3380	21000	-234.31
2200	-294.0	8500	-51.355	14800	-211.221	21100	297.125
2300	360.07	8600	101.215	14900	184.4506	21200	-188.80
2400	-312.2	8700	-40.235	15000	-186.672	21300	-20.538
2500	266.13	8800	60.3207	15100	120.1721	21400	111.999
2600	-94.77	8900	16.5930	15200	-171.940	21500	-64.634
2700	67.022	9000	-40.213	15300	139.2770	21600	86.9817
2800	-16.95	9100	43.3464	15400	-112.699	21700	-126.90
2900	40.613	9200	-17.046	15500	98.81825	21800	115.218
3000	-16.48	9300	83.1409	15600	-71.1666	21900	-236.49
3100	41.430	9400	-83.778	15700	15.28351	22000	225.481
3200	58.518	9500	148.306	15800	-21.1012	22100	-253.77
3300	-51.75	9600	-155.41	15900	54.25647	22200	275.618
3400	3.9449	9700	224.394	16000	-166.429	22300	-145.66
3500	27.971	9800	-146.45	16100	189.5156	22400	154.998
3600	74.111	9900	165.634	16200	-244.029	22500	-165.17
3700	-162.4	10000	-55.533	16300	224.1271	22600	103.662
3800	121.90	10100	21.9376	16400	-166.017	22700	-52.532
3900	-76.43	10200	-63.546	16500	177.6340	22800	-47.525
4000	-22.98	10300	161.568	16600	-121.951	22900	128.173
4100	96.511	10400	-220.69	16700	87.82277	23000	-54.910
4200	-80.24	10500	208.957	16800	-33.5085	23100	-110.45
4300	91.075	10600	-213.74	16900	-15.7863	23200	332.201
4400	-169.6	10700	303.710	17000	-79.4970	23300	-343.57
4500	190.21	10800	-250.44	17100	34.91056	23400	319.419
4600	-254.0	10900	298.953	17200	63.47601	23500	-274.02
4700	260.11	11000	-331.59	17300	-118.063	23600	221.795
4800	-290.1	11100	305.501	17400	113.3000	23700	-242.48
4900	268.35	11200	-353.50	17500	3.871122	23800	294.617
5000	-151.5	11300	406.846	17600	-92.1338	23900	-242.92
5100	142.32	11400	-372.14	17700	128.0973	24000	240.098
5200	-153.8	11500	308.390	17800	-136.380	24100	-255.43
5300	214.35	11600	-252.28	17900	155.4036	24200	238.043
5400	-311.0	11700	199.045	18000	-159.003	24300	-239.06
5500	550.61	11800	-237.02	18100	147.3679	24400	343.943
5600	-514.3	11900	147.105	18200	-66.7613	24500	-356.83
5700	426.30	12000	-86.620	18300	149.5479	24600	321.761
5800	-390.4	12100	76.5420	18400	-353.496	24700	-337.31
5900	526.71	12200	-144.74	18500	422.1357	24800	330.6571
6000	-518.1	12300	218.858	18600	-394.675	24900	-280.64

APPENDIX C2

AR(1) Coloured Noise generated with ϕ =-0.9 from a Random Noise of Mean zero & Variance 10,000

25000	308.701	31300	-77.681	37600	-7.2798	43900	67.18
25100	-188.49	31400	24.558	37700	-131.76	44000	-11.1
25200	126.132	31500	-6.1218	37800	155.970	44100	-30.2
25300	-66.562	31600	-70.246	37900	-134.59	44200	6.141
25400	159.40	31700	118.722	38000	150.315	44300	-106.
25500	-159.[illegible]	31800	-232.66	38100	-241.48	44400	41.67
25600	172.277	31900	177.[illegible]77	38200	320.693	44500	-50.7
25700	-40.431	32000	-146.51	38300	-307.98	44600	101.2
25800	-24.761	32100	76.1493	38400	279.804	44700	-99.[illegible]
25900	114.917	32200	-5.0876	38500	-177.65	44800	188.8
26000	-118.84	32300	-23.292	38600	107.673	44900	-172.
26100	211.928	32400	32.4204	38700	-196.16	45000	180.9
26200	-198.10	32500	1.99930	38800	229.948	45100	-218.
26300	293.070	32600	50.2652	38900	-182.09	45200	147.6
26400	-295.54	32700	-52.699	39000	104.572	45300	-160.
26500	257.249	32800	117.435	39100	-119.33	45400	208.4
26600	-252.06	32900	-35.751	39200	82.5302	45500	-145.
26700	318.007	33000	41.1532	39300	-79.993	45600	153.5
26800	-369.46	33100	-37.538	39400	129.199	45700	-131.
26900	375.098	33200	51.2909	39500	-121.70	45800	165.2
27000	-404.73	33300	45.9075	39600	38.6683	45900	-204.
27100	350.486	33400	-116.58	39700	-26.894	46000	174.0
27200	-354.90	33500	74.8242	39800	122.759	46100	-141.
27300	367.035	33600	-152.50	39900	-140.56	46200	58.39
27400	-338.37	33700	158.511	40000	158.722	46300	7.754
27500	355.377	33800	-174.34	40100	-95.214	46400	6.611
27600	-255.23	33900	97.7720	40200	41.4450	46500	-42.61
27700	192.949	34000	-175.07	40300	-46.419	46600	-15.84
27800	-76.522	34100	286.866	40400	51.7059	46700	62.127
27900	109.899	34200	-253.10	40500	-50.854	46800	-118.2
28000	-188.29	34300	210.122	40600	-43.928	46900	158.29
28100	187.025	34400	-252.65	40700	57.3788	47000	-208.5
28200	-139.52	34500	187.557	40800	4.69584	47100	103.45
28300	109.474	34600	-182.92	40900	-111.93	47200	-117.7
28400	-109.46	34700	180.018	41000	125.414	47300	18.504
28500	102.684	34800	-233.15	41100	-144.76	47400	26.073
28600	-3.5423	34900	107.625	41200	30.0304	47500	-102.5
28700	-185.43	35000	-143.89	41300	-25.264	47600	84.558
28800	105.046	35100	121.151	41400	73.2254	47700	-14.20
28900	-98.650	35200	-163.22	41500	-107.20	47800	-79.83
29000	22.0690	35300	179.629	41600	60.9615	47900	92.101
29100	-120.49	35400	-118.85	41700	-95.252	48000	-65.00
29200	160.516	35500	-50.777	41800	17.9096	48100	-9.839
29300	-239.25	35600	116.808	41900	21.8370	48200	106.20
29400	127.268	35700	-118.07	42000	-61.120	48300	-171.7
29500	-77.520	35800	136.333	42100	60.1304	48400	141.28
29600	110.248	35900	-152.03	42200	-28.842	48500	-66.66
29700	-55.520	36000	96.8593	42300	-20.807	48600	65.007
29800	101.518	36100	-69.575	42400	69.4918	48700	-126.9
29900	-84.696	36200	51.5699	42500	-134.00	48800	84.999
30000	16.8178	36300	-73.131	42600	120.286	48900	-119.9
30100	-27.263	36400	116.764	42700	-77.255	49000	176.74
30200	73.9062	36500	-210.91	42800	105.783	49100	-198.8
30300	-99.703	36600	177.341	42900	-192.70	49200	196.01
30400	79.7137	36700	-85.872	43000	142.870	49300	-248.3
30500	-57.201	36800	-45.686	43100	25.0996	49400	217.97
30600	-41.084	36900	-31.404	43200	-51.400	49500	-261.6
30700	-8.2678	37000	65.0301	43300	107.364	49600	202.50
30800	86.6427	37100	-47.725	43400	-157.96	49700	-73.5[illegible]
30900	-15[illegible].36	37200	81.6360	43500	173.844	49800	-56.29
31000	274.017	37300	-62.842	43600	-191.52	49900	59.054
31100	-205.60	37400	41.0368	43700	149.885	50000	-93.82
31200	99.5507	37500	-60.318	43800	-138.83		

APPENDIX D1

Linear Growth + AR(1) Coloured Noise with $\phi=-0.9$, from a Random Noise with Mean zero & Variance 10,000

		6100	635.965	12400	-45.715	18700	759.[illegible]
		6200	-355.75	12500	574.366	18800	-73.934
		6300	754.100	12600	13.8459	18900	639.789
100	155.165	6400	-358.61	12700	538.741	19000	-73.406
200	184.842	6500	687.882	12800	0.93784	19100	602.831
300	191.495	6600	-421.45	12900	422.015	19200	-1.4878
400	28.1329	6700	600.668	13000	-95.661	19300	370.372
500	332.892	6800	-301.56	13100	598.036	19400	199.243
600	-133.29	6900	540.636	13200	-114.67	19500	524.767
700	253.119	7000	-64.845	13300	372.666	19600	129.039
800	7.27846	7100	333.269	13400	314.707	19700	442.505
900	227.858	7200	112.373	13500	263.203	19800	216.789
1000	-135.49	7300	324.956	13600	225.635	19900	401.754
1100	343.246	7400	-80.826	13700	173.866	20000	163.891
1200	64.5678	7500	325.481	13800	247.174	20100	201.280
1300	441.824	7600	106.662	13900	52.2786	20200	276.447
1400	-107.14	7700	355.102	14000	410.483	20300	200.939
1500	269.250	7800	52.0267	14100	136.710	20400	428.262
1600	-25.197	7900	440.731	14200	431.643	20500	296.930
1700	398.445	8000	81.2867	14300	-79.456	20600	82.7686
1800	-206.20	8100	53.1733	14400	346.202	20700	462.312
1900	466.443	8200	324.172	14500	230.778	20800	16.0222
2000	-144.11	8300	44.5357	14600	85.6441	20900	500.977
2100	383.676	8400	284.567	14700	480.168	21000	37.9035
2200	-215.35	8500	162.760	14800	-2.8704	21100	668.274
2300	548.257	8600	326.429	14900	432.626	21200	178.355
2400	-179.34	8700	183.290	15000	52.3533	21300	166.070
2500	382.013	8800	266.880	15100	341.816	21400	491.014
2600	129.860	8900	255.784	15200	39.8426	21500	276.829
2700	182.849	9000	135.176	15300	384.569	21600	424.548
2800	141.469	9100	241.593	15400	152.554	21700	159.421
2900	187.955	9200	192.216	15500	354.792	21800	435.970
3000	128.327	9300	324.397	15600	199.567	21900	-5.0992
3100	191.693	9400	106.780	15700	242.122	22000	555.843
3200	256.573	9500	395.126	15800	234.618	22100	34.7426
3300	83.4052	9600	28.4897	15900	339.738	22200	631.264
3400	110.699	9700	481.155	16000	16.8939	22300	248.326
3500	185.721	9800	91.8286	16100	479.860	22400	497.181
3600	278.681	9900	390.415	16200	-29.045	22500	144.583
3700	-88.328	10000	210.651	16300	492.623	22600	401.326
3800	245.060	10100	207.199	16400	124.554	22700	303.955
3900	86.5296	10200	103.645	16500	464.224	22800	218.493
4000	56.5984	10300	435.070	16600	172.245	22900	516.704
4100	290.097	10400	-64.281	16700	342.675	23000	317.931
4200	67.7472	10500	424.045	16800	267.862	23100	114.731
4300	248.346	10600	-21.952	16900	224.900	23200	823.243
4400	-83.246	10700	588.982	17000	129.852	23300	-38.408
4500	361.984	10800	-23.688	17100	283.622	23400	662.505
4600	-162.50	10900	560.963	17200	401.942	23500	72.2732
4700	429.692	11000	-160.34	17300	116.352	23600	543.273
4800	-178.62	11100	524.640	17400	394.508	23700	67.7723
4900	423.512	11200	-191.17	17500	352.810	23800	686.317
5000	60.4272	11300	682.973	17600	126.670	23900	113.353
5100	298.761	11400	-158.78	17700	438.107	24000	596.901
5200	-17.716	11500	508.864	17800	129.965	24100	61.3373
5300	420.044	11600	-15.868	17900	459.010	24200	587.821
5400	-240.33	11700	400.532	18000	110.686	24300	89.4706
5500	887.897	11800	-54.760	18100	434.163	24400	777.159
5600	-369.21	11900	327.986	18200	262.431	24500	-41.273
5700	560.396	12000	168.631	18300	495.676	24600	670.578
5800	-235.34	12100	299.803	18400	-214.43	24700	-20.113
5900	805.812	12200	29.2443	18500	779.833	24800	699.256
6000	-385.88	12300	504.756	18600	-115.96	24900	[illegible]

APPENDIX D2

Linear Growth + AR(1) Coloured Noise with ϕ =-0.9, from a Random Noise with Mean Zero & Variance 10,000

25000	693.977	31300	346.116	37600	429.377	43900	568.872
25100	223.521	31400	528.943	37700	253.949	44000	512.259
25200	451.923	31500	465.579	37800	728.083	44100	466.526
25300	320.922	31600	297.875	37900	374.617	44200	535.116
25400	583.283	31700	626.597	38000	652.102	44300	368.783
25500	187.205	31800	121.900	38100	164.917	44400	549.816
25600	550.394	31900	578.873	38200	872.601	44500	486.153
25700	396.864	32000	285.771	38300	164.065	44600	685.640
25800	295.059	32100	462.613	38400	765.988	44700	442.105
25900	539.471	32200	462.740	38500	358.909	44800	804.641
26000	233.915	32300	383.990	38600	560.643	44900	376.742
26100	646.959	32400	467.211	38700	225.939	45000	749.685
26200	162.273	32500	451.130	38800	755.958	45100	295.343
26300	736.915	32600	514.561	38900	325.742	45200	667.339
26400	50.5494	32700	372.500	39000	556.620	45300	375.481
26500	619.846	32800	595.934	39100	356.625	45400	806.890
26600	103.549	32900	443.837	39200	559.679	45500	439.631
26700	750.006	33000	480.708	39300	411.043	45600	726.116
26800	-54.062	33100	396.636	39400	663.717	45700	430.328
26900	776.321	33200	498.626	39500	371.635	45800	755.706
27000	-76.528	33300	544.628	39600	488.804	45900	317.895
27100	715.599	33400	270.305	39700	477.300	46000	728.355
27200	-6.1552	33500	493.079	39800	689.182	46100	431.633
27300	775.524	33600	229.509	39900	340.179	46200	575.285
27400	33.6137	33700	613.236	40000	682.477	46300	612.516
27500	768.150	33800	245.671	40100	440.016	46400	580.935
27600	167.914	33900	500.193	40200	515.719	46500	498.806
27700	548.757	34000	209.368	40300	452.494	46600	514.659
27800	370.633	34100	818.265	40400	564.466	46700	662.452
27900	520.354	34200	195.655	40500	453.304	46800	408.734
28000	135.167	34300	644.532	40600	403.568	46900	763.289
28100	583.544	34400	151.744	40700	578.329	47000	317.928
28200	265.622	34500	608.853	40800	552.674	47100	618.530
28300	485.359	34600	256.655	40900	326.368	47200	438.387
28400	270.872	34700	640.503	41000	653.914	47300	533.177
28500	494.206	34800	169.929	41100	346.584	47400	629.507
28600	446.192	34900	490.596	41200	476.156	47500	419.752
28700	78.0749	35000	277.157	41300	490.567	47600	655.864
28800	454.819	35100	569.230	41400	622.963	47700	605.088
28900	290.998	35200	254.879	41500	381.651	47800	436.235
29000	370.434	35300	657.331	41600	554.640	47900	685.217
29100	205.843	35400	366.722	41700	396.072	48000	527.548
29200	590.786	35500	300.448	41800	491.636	48100	525.520
29300	92.9496	35600	623.281	41900	567.864	48200	754.332
29400	464.836	35700	332.771	42000	432.345	48300	360.324
29500	345.346	35800	617.203	42100	585.979	48400	716.707
29600	536.522	35900	289.774	42200	511.626	48500	559.582
29700	374.008	36000	532.435	42300	472.061	48600	654.901
29800	537.429	36100	405.805	42400	629.134	48700	414.278
29900	322.006	36200	508.641	42500	344.191	48800	653.608
30000	379.914	36300	374.338	42600	647.393	48900	440.030
30100	368.651	36400	617.631	42700	472.003	49000	813.452
30200	512.327	36500	185.179	42800	659.647	49100	365.710
30300	284.063	36600	637.275	42900	271.907	49200	799.842
30400	480.065	36700	433.259	43000	653.499	49300	296.345
30500	360.718	36800	341.754	43100	661.089	49400	808.411
30600	305.538	36900	390.874	43200	462.645	49500	289.352
30700	371.138	37000	561.973	43300	683.073	49600	776.308
30800	551.001	37100	432.755	43400	336.171	49700	596.785
30900	202.726	37200	581.965	43500	731.680	49800	459.180
31000	776.904	37300	419.617	43600	322.301	49900	663.702
31100	236.215	37400	506.856	43700	673.141	50000	473.707
31200	457.028	37500	401.165	43800	397.904		

APPENDIX_E1

AR(1) Coloured Noise generated with $\phi=0.5$ from a Random Noise of Mean zero & Variance 10,000

		6100	39.3005	12400	-27.227	18700	50.[illegible]
		6200	-44.246	12500	26.6381	18800	73.403
		6300	16.7811	12600	65.8711	18900	56.268
100	28.[illegible]	6400	4.36193	12700	91.7259	19000	2.4342
200	90.1590	6500	11.3148	12800	72.3455	19100	-17.41
300	14.[illegible]	6600	-66.743	12900	21.1947	19200	-23.25
400	76.2510	6700	-112.31	13000	-82.105	19300	-117.9
500	133.328	6800	-127.27	13100	-41.407	19400	-98.08
600	65.716	6900	-117.11	13200	-41.451	19500	16.736
700	-8.0844	7000	-27.669	13300	-130.81	19600	42.303
800	-0.[illegible]	7100	-42.505	13400	6.36092	19700	23.751
900	12.0333	7200	8.34676	13500	60.0876	19800	36.036
1000	-67.950	7300	53.1331	13600	37.4461	19900	36.240
1100	-47.710	7400	-35.013	13700	-31.639	20000	-4.005
1200	55.9416	7500	-83.097	13800	-65.782	20100	-127.5
1300	194.221	7600	-30.605	13900	-151.28	20200	-155.1
1400	169.308	7700	39.3921	14000	-112.72	20300	-165.1
1500	71.3837	7800	40.7816	14100	-49.526	20400	-84.85
1600	25.1292	7900	93.7672	14200	16.7837	20500	12.705
1700	85.6606	8000	125.336	14300	-73.466	20600	-102.6
1800	13.2079	8100	-45.599	14400	-165.20	20700	-99.15
1900	24.0126	8200	-43.893	14500	-79.715	20800	-139.0
2000	31.6282	8300	-42.571	14600	-139.37	20900	-124.4
2100	28.6069	8400	-48.706	14700	-53.066	21000	-123.5
2200	-43.736	8500	-4.1849	14800	-51.990	21100	-11.94
2300	22.5106	8600	56.2801	14900	-61.445	21200	108.84
2400	56.3716	8700	95.9796	15000	-62.870	21300	-57.82
2500	22.4822	8800	92.1742	15100	-91.338	21400	-27.19
2600	126.700	8900	117.195	15200	-134.16	21500	54.772
2700	112.355	9000	64.9849	15300	-123.50	21600	85.468
2800	91.5853	9100	21.7434	15400	-69.723	21700	26.831
2900	81.5133	9200	19.8339	15500	-36.032	21800	-0.302
3000	65.6999	9300	68.7105	15600	-4.4412	21900	-115.8
3100	59.1266	9400	50.7533	15700	-37.782	22000	-109.9
3200	119.284	9500	79.6992	15800	-50.068	22100	-101.3
3300	101.202	9600	44.6056	15900	-3.8042	22200	-28.68
3400	11.5587	9700	79.4423	16000	-90.332	22300	103.51
3500	0.27977	9800	116.797	16100	-67.987	22400	138.63
3600	86.3013	9900	110.730	16200	-91.006	22500	78.953
3700	-1.4334	10000	147.388	16300	-77.068	22600	-1.864
3800	-74.287	10100	85.3655	16400	-11.208	22700	16.167
3900	-42.668	10200	-21.726	16500	38.4153	22800	-54.07
4000	-103.37	10300	32.9377	16600	72.7923	22900	9.2895
4100	-42.139	10400	-22.630	16700	41.6933	23000	99.056
4200	2.02861	10500	-49.483	16800	55.4865	23100	-46.87
4300	15.6324	10600	-66.667	16900	9.40291	23200	114.27
4400	-70.097	10700	34.1869	17000	-97.342	23300	127.71
4500	-56.729	10800	74.8067	17100	-133.69	23400	77.643
4600	-103.26	10900	108.725	17200	-12.157	23500	59.798
4700	-75.350	11000	23.7732	17300	-15.540	23600	19.435
4800	-90.142	11100	-23.902	17400	-21.745	23700	-36.15
4900	-75.628	11200	-102.02	17500	83.6067	23800	28.216
5000	30.6203	11300	-31.931	17600	19.7152	23900	71.489
5100	57.9700	11400	-2.0356	17700	17.2105	24000	76.989
5200	10.2485	11500	-37.002	17800	6.93115	24100	21.433
5300	52.3193	11600	-26.293	17900	26.0748	24200	2.9286
5400	-53.884	11700	-44.877	18000	12.3743	24300	-18.05
5500	145.954	11800	-102.88	18100	4.44148	24400	94.451
5600	173.287	11900	-157.42	18200	62.3799	24500	71.305
5700	61.2723	12000	-92.402	18300	145.495	24600	27.035
5800	-2.7837	12100	-45.908	18400	-67.351	24700	-30.00
5900	134.584	12200	-101.08	18500	-43.936	24800	-15.36
6000	105.014	12300	-24.119	18600	-3.6616	24900	16.977

APPENDIX E2

AR(1) Coloured Noise generated with ϕ =0.5 from a Random Noise of Mean Zero & Variance 10,000

25000	[illegible]	31300	[illegible]	37600	[illegible]	43900	[illegible]
25100	[illegible]	31400	[illegible]	37700	[illegible]	44000	[illegible]
25200	[illegible]	31500	[illegible]	37800	[illegible]	44100	[illegible]
25300	71.772	31600	20.254	37900	25.847	44200	[illegible]
25400	[illegible]	31700	[illegible]	38000	[illegible]	44300	[illegible]
25500	[illegible]	31800	[illegible]	38100	[illegible]	44400	[illegible]
25600	[illegible]	31900	[illegible]	38200	[illegible]	44500	[illegible]
25700	154.48	32000	-56.83	38300	51.024	44600	-12.82
25800	[illegible]	32100	-76.73	38400	[illegible]	44700	25.469
25900	[illegible]	32200	-10.61	38500	93.127	44800	115.48
26000	73.860	32300	-2.923	38600	47.620	44900	120.83
26100	[illegible]	32400	0.5612	38700	-74.21	45000	106.14
26200	[illegible]	32500	32.293	38800	-36.81	45100	30.458
26300	[illegible]	32600	79.569	38900	28.302	45200	-43.39
26400	100.74	32700	63.737	39000	-12.66	45300	-66.21
26500	25.536	32800	95.958	39100	-46.66	45400	14.397
26600	-19.16	32900	145.44	39200	-55.10	45500	85.446
26700	48.963	33000	122.79	39300	-39.93	45600	102.80
26800	-14.87	33100	72.015	39400	34.157	45700	85.474
26900	-13.47	33200	50.308	39500	50.666	45800	100.67
27000	[illegible]	33300	111.98	39600	-24.27	45900	35.398
27100	-84.61	33400	34.324	39700	-31.36	46000	-6.184
27200	-100.7	33500	-43.02	39800	77.382	46100	8.6835
27300	-40.29	33600	-124.2	39900	72.049	46200	-42.05
27400	-13.72	33700	-96.86	40000	69.068	46300	6.7618
27500	29.541	33800	-79.85	40100	100.78	46400	47.956
27600	90.491	33900	-111.5	40200	47.126	46500	12.411
27700	43.600	34000	-169.4	40300	5.8901	46600	-49.87
27800	88.381	34100	-24.09	40400	10.278	46700	-5.919
27900	119.08	34200	47.561	40500	11.652	46800	-31.68
28000	2.0121	34300	22.479	40600	-68.13	46900	10.504
28100	-33.59	34400	-52.05	40700	-57.95	47000	-23.79
28200	-1.405	34500	-96.65	40800	27.578	47100	-107.3
28300	-10.08	34600	-89.48	40900	-44.67	47200	-116.5
28400	-29.86	34700	-44.23	41000	-42.12	47300	-149.4
28500	-27.99	34800	-79.17	41100	-39.31	47400	-76.22
28600	55.701	34900	-164.6	41200	-116.0	47500	-85.01
28700	-102.7	35000	-182.7	41300	-103.0	47600	-75.28
28800	-203.4	35100	-133.0	41400	-8.249	47700	17.749
28900	-165.3	35200	-122.3	41500	-6.620	47800	-30.53
29000	-162.0	35300	-57.35	41600	-40.70	47900	-25.77
29100	-213.2	35400	25.693	41700	-67.87	48000	17.518
29200	-125.1	35500	-96.13	41800	-109.1	48100	-28.89
29300	-135.7	35600	-55.23	41900	-50.46	48200	51.687
29400	-194.0	35700	-12.19	42000	-39.26	48300	14.643
29500	-121.0	35800	25.498	42100	-21.94	48400	-21.64
29600	-17.52	35900	8.8965	42200	22.245	48500	41.970
29700	53.246	36000	-36.04	42300	-5.188	48600	66.611
29800	100.31	36100	-18.95	42400	33.091	48700	-6.165
29900	88.503	36200	-8.159	42500	-13.14	48800	-50.66
30000	1.0664	36300	-25.74	42600	-27.25	48900	-74.25
30100	-40.32	36400	24.846	42700	18.558	49000	9.0532
30200	9.6208	36500	-50.85	42800	68.015	49100	12.023
30300	-2.803	36600	-78.96	42900	-18.59	49200	19.962
30400	-23.54	36700	16.215	43000	-70.68	49300	-35.39
30500	-7.824	36800	-58.19	43100	86.600	49400	-46.05
30600	-80.97	36900	-141.3	43200	103.70	49500	-78.89
30700	-127.0	37000	-79.22	43300	118.43	49600	-91.51
30800	-27.29	37100	-12.07	43400	46.364	49700	39.501
30900	-48.42	37200	45.892	43500	36.081	49800	-23.57
31000	61.833	37300	61.005	43600	7.0116	49900	-42.38
31100	130.11	37400	37.921	43700	-23.14	50000	-50.31
31200	24.519	37500	1.3263	43800	[illegible]		

APPENDIX_F1

Linear Growth + AR(1) Coloured Noise with $\phi = 0.5$ from a Random Noise of Mean Zero & Variance 10,000

		6100	[illegible]	12400	[illegible]	18700	[illegible]
		6200	67.9493	12500	[illegible]	18800	[illegible]
		6300	221.557	12600	[illegible]	18900	[illegible]
100	155.165	6400	171.707	12700	380.132	19000	[illegible]
200	359.928	6500	191.791	12800	332.315	19100	[illegible]
300	342.459	6600	41.6878	12900	245.299	19200	[illegible]
400	222.759	6700	-9.1916	13000	62.3002	19300	[illegible]
500	319.265	6800	-16.570	13100	199.983	19400	162.868
600	172.065	6900	9.86311	13200	178.271	19500	[illegible]
700	64.4966	7000	175.146	13300	12.4506	19600	370.940
800	112.305	7100	108.439	13400	311.553	19700	325.677
900	133.790	7200	212.039	13500	353.503	19800	358.196
1000	-21.997	7300	275.273	13600	288.117	19900	351.265
1100	52.4448	7400	89.9390	13700	168.610	20000	279.181
1200	240.361	7500	38.8930	13800	135.546	20100	64.7471
1300	457.195	7600	160.172	13900	-10.296	20200	67.3499
1400	350.912	7700	270.232	14000	100.980	20300	61.2674
1500	177.824	7800	243.026	14100	204.084	20400	217.924
1600	134.988	7900	343.751	14200	302.376	20500	367.637
1700	271.007	8000	381.320	14300	103.427	20600	107.281
1800	112.272	8100	45.4970	14400	-29.372	20700	165.69[illegible]
1900	161.141	8200	125.202	14500	173.363	20800	89.3850
2000	178.402	8300	127.811	14600	23.4616	20900	135.172
2100	163.433	8400	116.493	14700	213.642	21000	130.942
2200	30.3819	8500	204.253	14800	178.329	21100	342.301
2300	188.530	8600	300.086	14900	160.812	21200	522.317
2400	224.335	8700	349.368	15000	143.106	21300	155.018
2500	146.451	8800	325.967	15100	110.702	21400	287.496
2600	359.609	8900	376.270	15200	43.0203	21500	429.869
2700	287.789	9000	267.847	15300	82.6845	21600	452.832
2800	256.220	9100	210.383	15400	180.295	21700	329.636
2900	247.659	9200	226.976	15500	220.995	21800	305.346
3000	223.475	9300	321.323	15600	266.722	21900	99.8885
3100	219.255	9400	266.969	15700	190.683	22000	162.681
3200	337.125	9500	330.823	15800	183.078	22100	177.078
3300	277.705	9600	252.914	15900	276.847	22200	311.302
3400	117.506	9700	335.470	16000	93.3553	22300	530.087
3500	136.875	9800	391.965	16100	175.070	22400	539.344
3600	305.550	9900	365.350	16200	122.348	22500	412.396
3700	102.862	10000	438.721	16300	159.599	22600	287.208
3800	4.89817	10100	306.693	16400	278.738	22700	358.053
3900	97.6917	10200	132.887	16500	344.414	22800	218.280
4000	-30.220	10300	284.578	16600	388.597	22900	370.020
4100	113.101	10400	156.425	16700	315.812	23000	512.777
4200	170.414	10500	131.160	16800	356.789	23100	198.414
4300	177.614	10600	111.393	16900	264.716	23200	568.565
4400	10.5259	10700	310.519	17000	84.2648	23300	524.059
4500	74.5120	10800	343.840	17100	63.4284	23400	424.833
4600	-18.577	10900	391.265	17200	309.925	23500	414.484
4700	55.4657	11000	216.962	17300	250.725	23600	347.238
4800	15.9102	11100	165.467	17400	241.437	23700	261.018
4900	50.6156	11200	39.7473	17500	444.349	23800	408.240
5000	245.889	11300	207.548	17600	278.185	23900	462.687
5100	251.713	11400	233.985	17700	302.776	24000	455.176
5200	150.556	11500	135.634	17800	285.501	24100	318.924
5300	252.266	11600	192.247	17900	327.275	24200	339.575
5400	33.9515	11700	153.289	18000	293.981	24300	309.087
5500	459.718	11800	52.6722	18100	286.081	24400	532.080
5600	424.478	11900	-24.133	18200	400.175	24500	439.839
5700	202.111	12000	123.470	18300	532.930	24600	367.589
5800	131.723	12100	183.343	18400	95.1057	24700	280.38[illegible]
5900	420.844	12200	59.2944	18500	234.331	24800	331.217
6000	305.591	12300	229.934	18600	300.985	24900	389.851

APPENDIX_F2

Linear Growth + AR(1) Coloured Noise with $\phi = 0.5$
from a Random Noise of Mean Zero & Variance 10,000

25000	[illegible]	31300	378.150	37600	[illegible]	43900	[illegible]
25100	[illegible].537	31400	447.261	37700	465.760	44000	500.212
25200	[illegible].922	31500	590.646	37800	482.102	44100	[illegible]
25300	354.590	31600	419.172	37900	565.739	44200	427.202
25400	398.788	31700	515.[illegible]55	38000	612.164	44300	315.730
25500	306.879	31800	352.782	38100	395.284	44400	276.743
25600	[illegible]45.867	31900	274.158	38200	556.709	44500	395.116
25700	616.868	32000	353.199	38300	562.817	44600	571.034
25800	443.158	32100	304.315	38400	528.092	44700	596.959
25900	51[illegible].828	32200	439.932	38500	643.913	44800	751.317
26000	461.153	32300	425.196	38600	533.882	44900	722.979
26100	570.197	32400	429.369	38700	324.753	45000	694.216
26200	508.223	32500	488.841	38800	450.140	45100	558.725
26300	613.735	32600	565.313	38900	557.621	45200	453.830
26400	493.500	32700	515.716	39000	452.423	45300	444.208
26500	383.550	32800	584.861	39100	407.202	45400	607.530
26600	323.942	32900	665.502	39200	407.137	45500	707.276
26700	475.294	33000	602.253	39300	440.396	45600	709.938
26800	325.579	33100	517.631	39400	574.652	45700	670.817
26900	357.438	33200	500.279	39500	574.209	45800	708.495
27000	266.335	33300	627.691	39600	426.331	45900	579.649
27100	245.431	33400	454.396	39700	446.439	46000	530.995
27200	225.787	33500	344.085	39800	656.823	46100	578.486
27300	347.869	33600	225.651	39900	599.744	46200	476.971
27400	368.495	33700	313.967	40000	597.695	46300	592.609
27500	444.176	33800	334.647	40100	660.214	46400	649.936
27600	540.519	33900	267.962	40200	546.028	46500	565.644
27700	426.175	34000	173.148	40300	492.730	46600	464.602
27800	532.751	34100	474.030	40400	520.489	46700	576.130
27900	572.303	34200	546.188	40500	522.177	46800	508.941
28000	338.697	34300	468.210	40600	371.710	46900	601.015
28100	324.173	34400	339.425	40700	427.044	47000	518.516
28200	401.687	34500	288.882	40800	595.108	47100	376.557
28300	372.043	34600	322.871	40900	411.714	47200	396.887
28400	339.443	34700	405.980	41000	449.359	47300	339.362
28500	352.636	34800	320.760	41100	454.398	47400	492.828
28600	511.014	34900	175.397	41200	308.844	47500	444.575
28700	175.577	35000	179.398	41300	368.096	47600	467.589
28800	55.7454	35100	281.835	41400	542.269	47700	639.979
28900	172.862	35200	280.642	41500	504.386	47800	508.564
29000	162.752	35300	399.385	41600	440.177	47900	539.830
29100	64.8885	35400	528.512	41700	404.752	48000	620.501
29200	254.343	35500	262.454	41800	339.501	48100	514.636
29300	195.961	35600	394.394	41900	469.378	48200	688.452
29400	91.0083	35700	458.522	42000	465.428	48300	583.91[illegible]
29500	255.193	35800	511.710	42100	494.106	48400	532.706
29600	419.373	35900	464.732	42200	570.930	48500	670.001
29700	508.346	36000	388.112	42300	500.570	48600	689.586
29800	567.233	36100	441.201	42400	586.077	48700	542.264
29900	525.388	36200	454.988	42500	482.813	48800	491.273
30000	366.545	36300	418.689	42600	477.522	48900	467.18[illegible]
30100	327.946	36400	522.123	42700	571.322	49000	635.925
30200	441.594	36500	357.839	42800	645.638	49100	605.713
30300	397.035	36600	339.070	42900	461.190	49200	620.448
30400	364.304	36700	532.164	43000	401.920	49300	513.307
30500	404.129	36800	350.406	43100	723.185	49400	518.517
30600	259.614	36900	226.966	43200	687.231	49500	461.97[illegible]
30700	204.236	37000	381.678	43300	708.844	49600	453.39[illegible]
30800	415.764	37100	481.978	43400	548.356	49700	[illegible]
30900	332.233	37200	562.806	43500	580.734	49800	[illegible]
31000	550.942	37300	572.584	43600	531.690	49900	[illegible]
31100	632.504	37400	516.441	43700	488.526	50000	518.81[illegible]
31200	404.014	37500	459.893	43800	506.685		

APPENDIX G

Availability of Electricity (G.B) 1901-1980

Year	Y	Year	Y
1901	0.294	1941	31.01
1902	0.329	1942	33.841
1903	0.364	1943	35.096
1904	0.504	1944	36.585
1905	0.649	1945	35.706
1906	0.767	1946	39.178
1907	1.015	1947	40.369
1908	1.153	1948	43.815
1909	1.337	1949	46.241
1910	1.426	1950	51.877
1911	1.509	1951	56.658
1912	1.684	1952	58.816
1913	1.78	1953	61.955
1914	2.189	1954	68.921
1915	2.247	1955	75.648
1916	2.629	1956	82.191
1917	3.08	1957	85.999
1918	3.502	1958	93.441
1919	3.931	1959	100.572
1920	4.065	1960	114.417
1921	3.664	1961	122.823
1922	4.376	1962	136.015
1923	5.177	1963	147.102
1924	5.899	1964	154.534
1925	6.492	1965	166.051
1926	6.788	1966	171.999
1927	8.102	1967	177.309
1928	8.979	1968	190.166
1929	9.979	1969	202.593
1930	10.475	1970	211.461
1931	10.913	1971	218.053
1932	11.701	1972	224.531
1933	12.964	1973	239.651
1934	14.769	1974	231.726
1935	16.817	1975	232.431
1936	19.404	1976	234.458
1937	21.841	1977	241.174
1938	23.338	1978	246.642
1939	25.189	1979	258.045
1940	27.327	1980	246.053

Unit:- 1000 GWH

Source:- Digest of Energy Statistics (1934-1982)
Ministry of Power,U.K.

U.K. Chemical Industries Indices (1952-1982)

Year	1	2	3	4
1952	28.60	27.90	26.50	28.50
1953	31.10	31.10	30.30	33.00
1954	34.20	35.00	33.30	35.00
1955	36.60	36.30	34.50	37.00
1956	39.30	39.80	36.90	37.80
1957	39.90	41.10	39.00	40.80
1958	42.40	42.10	39.30	42.70
1959	45.00	47.00	44.20	47.70
1960	50.70	51.90	49.10	51.10
1961	53.40	52.90	48.80	51.20
1962	53.20	54.40	51.90	53.20
1963	55.30	58.50	55.70	58.30
1964	62.40	64.30	60.60	63.50
1965	67.90	69.40	64.40	66.20
1966	71.80	71.70	68.80	70.70
1967	75.30	76.30	71.00	74.10
1968	80.10	81.60	77.50	81.10
1969	84.90	86.70	82.80	85.20
1970	88.00	92.20	88.80	90.80
1971	92.60	93.00	87.40	94.10
1972	93.20	99.30	92.80	100.60
1973	106.70	107.20	105.50	109.70
1974	112.60	114.60	108.20	106.20
1975	103.20	99.70	94.60	102.40
1976	109.30	111.50	108.60	115.40
1977	118.70	115.80	109.30	112.80
1978	113.20	117.20	112.70	117.00
1979	115.90	121.90	114.50	117.40
1980	119.00	109.60	99.10	103.00
1981	104.90	106.60	104.10	106.60
1982	107.30	106.20	101.10	-

BASE:- 1975 = 100

Source:- I.C.I (U.K.)

APPENDIX I

Population of U.S.A (1900-1981)

Year	Y	Year	Y
1900	76.1	1941	133.9
1901	77.6	1942	135.4
1902	79.2	1943	137.3
1903	80.6	1944	138.9
1904	82.2	1945	140.5
1905	83.8	1946	141.9
1906	85.4	1947	144.7
1907	87.0	1948	147.2
1908	88.7	1949	149.8
1909	90.5	1950	152.3
1910	92.4	1951	154.9
1911	93.9	1952	157.6
1912	95.3	1953	160.2
1913	97.2	1954	163.0
1914	99.1	1955	165.9
1915	100.5	1956	168.9
1916	102.0	1957	172.0
1917	103.3	1958	174.9
1918	103.2	1959	177.8
1919	104.5	1960	180.7
1920	106.5	1961	183.7
1921	108.5	1962	186.6
1922	110.1	1963	189.2
1923	112.0	1964	191.9
1924	114.1	1965	194.3
1925	115.8	1966	196.6
1926	117.4	1967	198.7
1927	119.0	1968	200.7
1928	120.5	1969	202.7
1929	121.8	1970	205.1
1930	123.1	1971	207.7
1931	124.1	1972	209.9
1932	124.8	1973	211.9
1933	126.6	1974	213.9
1934	126.4	1975	216.0
1935	127.3	1976	218.0
1936	128.1	1977	220.2
1937	128.8	1978	222.6
1938	129.8	1979	225.1
1939	130.9	1980	227.7
1940	132.6	1981	229.8

Unit:- Millions

Source:- Statistical Abstracts of the United States(1982)
U.S. Department of Commerce,
Bureau of the Census

APPENDIX J

Austrian Disposable Personal Income (1954-1979)

Year	1	2	3	4
1954	17.62	18.86	20.58	23.51
1955	19.31	21.42	23.93	26.91
1956	21.98	24.00	25.96	27.66
1957	23.34	25.80	27.23	29.87
1958	24.19	26.70	28.47	31.65
1959	25.41	27.94	29.61	32.98
1960	27.40	29.40	32.04	35.37
1961	29.52	31.51	33.94	38.06
1962	30.70	32.87	34.75	38.76
1963	31.22	35.12	36.44	41.44
1964	33.72	36.77	37.71	43.61
1965	34.96	38.47	39.15	45.05
1966	36.57	40.68	41.10	46.43
1967	37.88	41.79	42.36	47.74
1968	40.61	43.93	43.92	49.01
1969	42.08	46.31	45.94	51.38
1970	44.53	48.65	47.99	54.08
1971	48.25	52.21	51.51	58.11
1972	50.90	55.29	54.00	60.62
1973	53.14	58.74	57.69	65.01
1974	57.71	62.36	59.88	66.60
1975	57.78	63.93	61.07	68.43
1976	60.37	66.05	63.47	72.19
1977	64.69	69.50	64.46	73.80
1978	66.02	70.48	65.35	74.70
1979	68.33	72.94	67.74	77.31

Unit:- billions of Schillings

Source:- Austrian Institute of Economic Research,
Vienna, Austria.

APPENDIX K

$\underline{f}$ Vectors and $\underline{G}$ Transition Matrices

The $\underline{f}$ vectors considered are (1xn) row vectors of some known functions of independent variables or constants. By definition the elements of these vectors are functions of time, generally described by constants, polynomials and trigonometric functions as in Brown (1962).

Time series may be defined such that for some constant n square matrix $\underline{G}$ and integer k

$$\underline{f}_{t+k} = \underline{f}_t \, \underline{G}^k$$

where the eigenvalues of the matrix $\underline{G}$ determine the form of forecast function. It is then usual to adopt a moving parameterisation as in Brown (1962), so that at time t given data $D_t = y_t, y_{t-1}, \ldots, y_1$ the current value of $\underline{f}_t$ is always $\underline{f}_0$. For simplicity we write $\underline{f}_0 = \underline{f}$.

For example, suppose at time t we have the linear growth representation with $\underline{f}_t = (1 \quad t)$ and parameter vector $\underline{\theta}' = (\theta_1 \quad \theta_2)$ and local model

$$Y_{t+k} = \underline{f}_{t+k} \, \underline{\theta} + \delta_{t+k} \quad \text{where} \quad \delta_{t+k} \text{ is white noise.}$$

The interpretation is that θ_1 is the intercept at t = 0 and θ_2 is the incremental growth per unit time. Then

$$(1 \quad t) = (1 \quad 0) \, \underline{G}^t$$

is satisfied for all integer $t \geq 0$ if and only if $\underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

Hence for time t the local model can be written as

$$\begin{aligned} Y_{t+k} &= \underline{f}_{t+k} \, \underline{G}^{-t} \, \underline{G}^t \, \underline{\theta} + \delta_{t+k} \\ &= \underline{f}_k \, \underline{\Omega} + \delta_{t+k} \end{aligned}$$

where $\underline{\Omega} = \underline{G}^t \, \underline{\theta}$ is the parameter vector at time t.

Akaike H (1971)
Autoregressive Model Fitting for Control.
Annals Inst. Stl. Math., 23 :pp163-180

(1974)
Markovian Representation of Stochastic Processes and its Applications to the Analysis of Autoregressive Moving Processes.
Annals Inst. Stl. Math., 26 :pp363-387

(1980)
The Interpretation of Improper Prior Distribution as Limits of Data Dependent Proper Prior Distributions.
J.R.S.S.(B), 42,1 :pp46-52

Akram M - Harrison P J (1983)
Some Applications of the Generalised Exponentially Weighted Regression (G.E.W.R).
International Time Series and Forecasting Meeting held at Nottingham (U.K.), April 11-15, 1983.
(to appear)

Ameen J R M - Harrison P J (1983)
Normal Discount Bayesian Models.
International Meeting on Bayesian Statistics,held at Valencia,(Spain), Sept. 6-10, 1983.
(to appear)

Andel J (1981)
An Autoregressive Representation of ARMA-Processes.
Proc. of Second Symposium on Mathl. Statistics.
Bad Tatzmannsdrof (Austria), June 14-20 :pp13-21.

Anderson B D O - Richard J C (1982)
Exponential Convergence of Adaptive and Control Algorithms.
Automatica, 18,1 : pp1-13.

Astrom K J (1981)
Theory and Applications of Adaptive Control.
IFAC - Control Science and Technology,KYOTO(Japan) pp 737-748.

Bartlett M S (1946)
On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series.
J.R.S.S. (B), 8 : pp27-41

Beck J V (1974)
Parameter Estimation with Cumulative Errors.
Technometrics, 16,1 : pp85-92

Bennett R J (1976)
Non-Stationary Parameter Estimation for Small Sample Situtations : A Comparison of Methods.
Int. J. Systems Sc., 7,3 : pp257-275

Box G E P - Jenkins G M - MacGregor J F (1974)
Some Recent Advances in Forecasting and Control (ii).
J.R.S.S. (Applied), 23, 2 : pp158-179

Box G E P - Jenkins G M (1970, 1976)
Time Series Analysis, Forecasting and Control.
Holden-Day, U.S.A.

Brown R G (1959, 1962)
Smoothing, Forecasting and Prediction of Discrete Time Series. Prentice Hall.

Brown R G - Meyer R F (1961)
The Fundamental Theorem of Exponential Smoothing.
Op. Res., 9 :pp673-685.

Bunn D W (1977)
Comparative Evaluation of the Outperformance and Minimum Variance Procedures for the Linear Synthesis of Forecasts.
Op. Res. Q.,28,3 : pp653-662.

Cains P E - Rissaner J (1974)
Maximum Likelihood Estimate of Parameters in Multivariate Gaussian Stochastic Processes.
I.E.E.E Trans. on Information Theory, IT20 :pp102-104.

Cantarelis N - Johnston F R (1982)
On-Line Variance Estimation for the Staedy State Bayesian Model.
J.Time Series Analysis, 3,4 : pp225-233.

Chow G C (1973)
Multiperiod Predictions for Stochastic Difference Equations by Bayesian Methods.
Econometrica, 41,1 : pp109-118.

Clarke B R - Godolphin E J (1982)
Comparative Power Studies for Goodness of Fit Tests of Time Series Models.
J. Time Series, 3,3 : pp141-151.

Cogger K O (1973)
Extensions of the Fundamental Theorem of Exponential Smoothing.
Mang. Sc., 19,5 : pp547-554.

(1974)
The Optimality of General Order Exponential Smoothing.
Op. Res., 22, :pp858-867

Cox D R (1961)
Prediction by Exponentially Weighted Moving Average and Related Methods.
J.R.S.S(B), 23 : pp414-422.

(1981)
Statistical Analysis of Time Series: Some Recent Devlopments.
Scand. J. Statist., 8 : pp93-115

DeGroot M H (1970)
Optimal Statistical Decisions.
McGraw-Hill.

D'Esopo D A (1961)
A Note on Forecasting by the Exponential Smoothing Operator.
Op. Res. ,9,5 :pp686-687

De Finetti B (1964)
Bayesianism : Its Unifying Role for both the Foundations and Applications of Statistics.
Int. Stl. Rev., 42,2 : pp 117-130

Dobbie J M (1963)
Forecasting Periodic Trends by Exponential Smoothing.
Op. Res., 11 : pp908-918

Durbin J (1959)
Efficient Estimation of Parameters in Moving Averages.
Biometrika, 46, part(i,ii), : pp306-316.

(1983)
Extensions of the Brown and Holt Winters Forecasting Systems and their relation to Box-Jenkins Models.
Time Series Analysis: Theory & Practice,3
North-Holland Pub. Co.(Ed. O D Anderson), pp7-18

Fukao T (1965)
System Identification by Bayesian Learning Process.
Proc. IFAC, Tokyo Symposium, Japan. pp137-46.

Friedland B (1981)
Discrete Time Equivalent of Continuous Time Filtering Processes.
J. Dynamic Systems, Measurement and Control; 103, pp417-419.

Fuller W A - Hasza D P (1981)
Properties of Predictors for Autoregressive Time Series.
J.A.S.A, 76,373 : pp155-161.

Gardner G - Harvey A C - Philips G D A (1980)
An Algorithm for Exact Maximum Likelihood Estimate of ARMA by means of Kalman Filtering.
J.R.S.S. (Applied),

Gathercole R B (1982)
A Note for Bayesian Forecasters.
Time Series Analysis: Theory & Practice,1
North Holland Pub. Co. (Ed. O D Anderson), pp487-495.

Genin Y (1970)
Further Comments on the Derivation of Kalman Filters.
AGARDOgraph, 139, Leondes ed., NATO Advanced Studies.
pp55-63.

Geurts M D - Ibrahim I B (1978)
Comparing the Box-Jenkins Approach with the Exponentially Smoothed Forecasting Model Application to Hawaii Tourists.
J. Mark.Res.. 12, :pp182-188

Gilchrist W G (1967)
Methods of Estimation involving Discounting.
J.R.S.S.(B),29, 2 :pp355-369.

Godolphin E J (1977)
A Direct Representation for the Maximum Likelihood Estimator of a Gaussian Moving Average Process.
Biometrika, 64, 2 : pp375-484

(1980)
A Method for Testing the Order of an ARMA-Process.
Biometrika, 67, 3 :pp699-703.

Godolphin E J - Harrison P J (1975)
Equivalence Theorems for Polynomial Projecting Predictors.
J.R.S.S.(B),37, : pp205-215.

Godolphin E J - Stone J M (1980)
On the Structural Representation for Polynomial-Projecting Predictor Models based on Kalman Filter.
J.R.S.S.(B), 42, 1 : pp35-45.

Goodman M L (1974)
A New Look at Higher Order Exponential Smoothing for Forecasting.
Op. Res., 22 :pp880-888.

Goodwin G C - Ramadge P J - Cains P E (1981)
A Globally Convergence Adaptive Predictor.
Automatica, 17, 1 : pp135-140.

Granger C W J - Newbold P (1976)
Forecasting Transformed Series.
J.R.S.S.(B), : pp189-203.

Grimble M J (1982)
Weighted Minimum Variance Self Tuning Control.
Int. J. Control, 36, 4 : pp597-609.

Groff G K (1973)
Empirical Comparison of Models for Short Range Forecng.
Mang. Sc., 20, 1 : pp22-31.

Harris T J - McGregor J F - Wright J D (1980)
Self Tuning and Adaptive Controllers: An Application to Catalytic Control.
Technometrics, 22, 2 : pp153-164.

Harrison P J (1965)
Short Term Sales Forecasting.
J.R.S.S.(C), 14, 2 : pp102-139.

(1967)
Exponential Smoothing and Short Term Sales Forecasts.
Mang. Sc., 13, 11 : pp821-842.

Harrison P J - Scott F A (1965)
A Development System for use in Short Term Sales Forecasting Investigations.
Paper presented to the O.R. Society Annual Conf. Shrivenham.

Harrison P J - Stevens C F (1971)
A Bayesian Approach to Short Term Forecasting.
Op. Res. Q., 22 : pp341-362.

(1976)
Bayesian Forecasting with Discussion.
J.R.S.S.(B), 38, 3 : pp205-247.

Harrison P J - Pearce (1972)
The use of Trend Curves as an aid to Market Forecasting.
Ind. Markt. Mang., 2 pp149-170.

Harrison P J - Gazzard T - Leonard T (1977)
An Application of Multivariate Hierarchical Forecasting.
Paper presented to the Royal Stl. Soc.:Industrial Applications Group(Internal Report, Statistics Dept., Warwick University)

Harrison P J - Akram M (1983)
Generalised Exponentially Weighted Regression and Parsimonious Dynamic Linear Modelling.
Time Series Analysis: Theory and Practice 3. Proc. of the Intl. Conf. held in Valencia, Spain, May 1982.
North-Holland Publishing Co. pp19-38

(1983)
Authors reply to McKenzie's Comments. pp41-42

Harrison P J - Johnston F R (1983)
A Regression Method with non stationary Parameters.
Internal Report No. 35, Statistics Department, Warwick University, Coventry, U.K.

Ho Y C - Lee R C K (1964)
A Bayesian Approach to problems in Stochastic Estimation and Control. Trans. I.E.E.E.,AC9 pp333-339

Holt C C (1957)
Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages.
Tech. Res. Memo No. 52, Carnegie Institute, U.S.A.

Jazwinski A H (1969)
Adaptive Filtering.
Automatica, 5 : pp475-485.

Jones R H (1966)
Exponential Smoothing for Multivariate Time Series.
J.R.S.S.(B),1 : pp241-251.

Kailath T (1974)
A view of three decades of Linear Filtering Theory.
I.E.E.E. Trans. Inf. Theory. IT-20(2), pp146-181.

Kalman R E (1960)
On the General Theory of Control Systems.
First IFAC Confernece on Automatic Control, Moscow Part 4 :pp2020-2030.

(1960)
A New Approach to Linear Filtering and Prediction Problems.
ASME J. Basic Eng., 82 :pp35-45.

Kalman R E - Koepcke R W (1958)
Optimal Synthesis of Linear Sampling Control Systems using Generalised Preference Indices.
Trans. ASME, 80 :pp1820-1826.

Kendall M G (1944)
On Autoregressive Time Series.
Biometrika, 33,3 :pp105-122.

(1961)
Studies in the History of Probability and Statistics.
Biometrika, 48, 1,2

Key P B - Godolphin E J (1981)
On the Bayesian Steady Forecasting Models.
J.R.S.S.(B), 43,1 : pp92-96.

Kortum W (1979)
Computational Techniques in Optimal State Estimation: A Tutorial Review.
J. Dynamic Systems, Measurement and Control,101,pp99-107

Ledolter J (1969)
A Recursive Approach To Parameter Estimation in Regression and Time Series Models.
Commun. Statist. Theor. Math., A8(12) :pp1227-1245

(1981)
Recursive Estimation and Adaptive Forecasting in ARIMA Models with Time Varying Coefficients.
Applied Time Series Analysis (ii), Ed. Findley D F
Academic Press. :pp449-471.

Ledolter J - Box G E P - Tio G C (1976)
Various Aspects of Parameter Changes in ARIMA Models.
Technical Report No. 449, Stat. Dept., Wisconsin Univ.

Lindley D V (1975)
The Future of Statistics: A Bayesian 21st Century.
Proc. of the Conf. on Directions for Mathl. Statistics
Univ. of Alberta, Edmonton, Canada, Aug. 12-16, 1974.

Lindley D V - Smith A F M (1972)
Bayes Estimates for the Linear Model.
J.R.S.S.(B), 1 :pp1-41.

Ljung L (1978)
Convergence Analysis of Parameteric Identification Methods.
I.E.E.E. Trans. Auto. Cont., AC23, 5 :pp770-783.

Makridakis S - Hibon M (1979)
Accuracy of Forecasting: An Empirical Investigation.
J.R.S.S.(A), 142,2 :pp97-145.

McGregor J F (1972)
State Variable Approach to Time Series Representation and Forecasting.
Technical Report, Statistics Department, University of Wisconsin, Madison, U.S.A.

McKenzie E (1974)
A Comparison of Standard Forecasting System with Box-Jenkins Approach.
Statistician, 23 :pp107-116

(1976)
An Analysis of General Exponential Smoothing.
Op. Res., 24 :pp131-140.

(1976)
A Comparison of Some Standard Seasonal Forecasting Sys.
Statisticain, 25, 1 :pp3-14.

McKenzie E - Peterson M (1978)
A Note on a Matrix Inversion in Multiple Exponential Smoothing.
J.R.S.S.(B), 40, 2 :pp182-183.

McLeod I (1975)
Derivation of the Theoretical Autocovariance Function of Autoregressive Moving Average Time Series.
J.R.S.S.(Applied), 24, 2 :pp255-256.

Melard G (1980)
Some Links between the Harrison-Stevens and Box-Jenkins Methods for Univariate Time Series.
Time Series. Ed. O D Anderson. North Holland Pub. Co.
pp319-323.

Meyer R F (1963)
An Adaptive Method for Routine Short Term Forecasting
IFORS, Oslo, Norway, July 1963. :pp882-894.

Morrison G W - Pike D H (1977)
Kalman Filtering Applied to Statistical Forecasting.
Mang. Sc., 23, 7 : pp768-774.

Muth J F (1960)
Optimal Properties of Exponentially Weighted Forecasts.
J.A.S.A., 55. : pp299-306.

Newbold P (1981)
Some Recent Developments in Time Series Analysis.
Int. Stl. Rev., 49 :pp53-66.

Newbold P - Granger C W J (1974)
Experience with Forecasting Univariate Time Series and the Combination of Forecasts.
J.R.S.S.(A), 137, 2 :pp131-164.

Nuri W A (1980)
Estimation in Regression Models with Errors following an Autoregressive Moving Average Scheme.
Time Series Analysis. Ed. O D Anderson. North Holland Pub. Co. :pp325-331.

Pack C D - Whitaker B A (1982)
Kalman Filter Models for Network Forecasting.
The Bell System Tech. J. , 61,1 :pp1-14.

Pandit S M - Wu S M (1974)
Exponential Smoothing as a special case of a Linear Stochastic System.
Op. Res., 22 :pp868-879.

Parzen E (1974)
Some Recent Advances in Time Series Modelling.
I.E.E.E.Trans., AC19, 6 :pp723-730.

Peterka V (1981)
Bayesian System Identification.
Automatica, 17, 1 :pp41-53.

Plackett R L (1950)
Some Theorems in Least Squares.
Biometrika, 37, 1&2 :pp149-157.

Pleszczynska E - Dabrowska D (1980)
On Partial Observability in Statistical Models.
Math. Operationsforsch Statist., Ser. Statistics, 11, 1 :pp49-59.

Prins J (1982)
Interactive Comparison of Forecasting Methods.
Time Series Analysis: Theory and Practice, 1
North Holland Pub. Co. Ed. O D Anderson. pp443-461.

Revankar N S (1980)
Analysis of Regressions Containing Serially Correlated and Serially Uncorrelated Error Coefficients.
Int. Econ. Review, 21, 1 :pp185-199.

Roberts S A (1982)
The Formulation of Forecasting Models and Box-Jenkins Modelling.
(to appear)

Roberts S A - Harrison P J (1982)
Parsimonious Modelling & Forecasting Seasonal Time Serie
Eur. J. Op. Res. (to appear)

Sarris A H (1973)
Kalman Filter Models: A Bayesian Approach to Estimation of Time Varying Regression Coefficients.
Annals of Econ. and Social Measurement, 2, 4 :pp501-523

Singer R A - Frost P A (1969)
On the Relative Performance of the Kalman and Wiener Filters.
I.E.E.E. Trans. on Automatic Control. :pp390-394

Slutzky E (1927)
The Estimation of Random Cases as the Source of Cyclic Processes.
English Translation (1937):Econometrica,5 :pp105-106.

Smith J Q (1979)
A Generalisation of the Bayesian Staedy Forecasting Models.
J.R.S.S.(B), 41, 3 :pp375-87.

Stigler M S (1983)
Who Discovered Bayes Theorem ?
The American Statistician, 37, 4, 1 :pp290-296

Stoica P - Soderstrom T (1982)
On the Parsimony Principle.
Intl. J. Control, 36, 3 :pp409-418.

Stoodley K D C - Lewis T - Stainton C L S (1980)
Applied Statistical Techniques.
Elis Horwood Series, J. Wiley. :PP235-275.

Szelag C R (1982)
A Short Term Forecasting Algorithm for Trunk Demand Servicing.
The Bell System Tech. J., 61, 1 :pp67-96.

Thury G (1983)
Applied Multiple Time Series Analysis: A Case Study of Austria.
International Time Series and Forecasting Meeting held at Nottingham (U.K.), April 11-15,1983.

Tee L H - Wu S M (1972)
An Application of Stochastic and Dynamic Models for the Control of a Paper Making Process.
Technometrics, 14, 2 :pp481-496.

Walker G (1931)
On Periodicities in Series of Related Terms.
Proc. Royal Soc., London. :pp518-532.

Ward D H (1963)
Comparison of Different Systems of Exponentially Weighted Prediction.
Statistician, 13, 3 :pp137-191.

West M (1981)
Robust Sequential Approximate Bayesian Estimation.
J.R.S.S.(B)., 43, 2 :pp157-166.

(1982)
Aspects of Recursive Bayesian Estimation.
Ph.D Thesis, University of Nottingham(U.K.).

Whittle P (1963)
Prediction and Regulation by Linear Least Squares Methods.
English Universities Press.

Wichern W D (1973)
The Behaviour of Sample Autocorrelation Function for IMA-Process.
Biometrika, 60, 2 :pp235-239

Wichern W D - Hsu D A (1976)
Changes of Variance in First Order Autoregressive Time Series Models with an Application.
J.R.S.S.(Applied), 25, 3 :pp248-256.

Wilson D R - Newman L M (1981)
Model Selection Via Accuracy Criteria: Box - Jenkins Vs EWMA Methods.
Internal Report, College of Business Administration Pennsylvania State University, U.S.A.

Winter P R (1960)
Forecasting Sales by Exponentially Weighted Moving Average.
Mang. Sc., 6 :pp324-342.

Wold H O (1938)
A Study in the Analysis of Stationary Time Series.
Almquist and Wiksell, Uppsala, Sweden.

Yule U G (1927)
On the Methods of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers.
Philosophical Trans., A226. :pp267-298.

Whittle P (1963)
Prediction and Regulation by Linear Least Squares Methods.
English Universities Press.

Wichern W D (1973)
The Behaviour of Sample Autocorrelation Function for IMA-Process.
Biometrika, 60, 2 :pp235-239

Wichern W D - Hsu D A (1976)
Changes of Variance in First Order Autoregressive Time Series Models with an Application.
J.R.S.S.(Applied), 25, 3 :pp248-256.

Wilson D R - Newman L M (1981)
Model Selection Via Accuracy Criteria: Box - Jenkins Vs EWMA Methods.
Internal Report, College of Business Administration Pennsylvania State University, U.S.A.

Winter P R (1960)
Forecasting Sales by Exponentially Weighted Moving Average.
Mang. Sc., 6 :pp324-342.

Wold H O (1938)
A Study in the Analysis of Stationary Time Series.
Almquist and Wiksell, Uppsala, Sweden.

Yule U G (1927)
On the Methods of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers.
Philosophical Trans., A226. :pp267-298.